

Attorney Docket No. 1095.1308

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:

Kazunari TANAKA et al.

Application No.: Unassigned

Group Art Unit: Unassigned

Filed: March 30, 2004

Examiner: Unassigned

For: RELATION CHART-CREATING PROGRAM, RELATION CHART-CREATING METHOD,
AND RELATION CHART-CREATING APPARATUS

**SUBMISSION OF CERTIFIED COPY OF PRIOR FOREIGN
APPLICATION IN ACCORDANCE
WITH THE REQUIREMENTS OF 37 C.F.R. § 1.55**

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Sir:

In accordance with the provisions of 37 C.F.R. § 1.55, the applicant(s) submit(s) herewith
a certified copy of the following foreign application:

Japanese Patent Application No(s). 2003-353928

Filed: October 14, 2003

It is respectfully requested that the applicant(s) be given the benefit of the foreign filing
date(s) as evidenced by the certified papers attached hereto, in accordance with the
requirements of 35 U.S.C. § 119.

Respectfully submitted,

STAAS & HALSEY LLP

By:

William F. Herbert
Registration No. 31,024

Date: March 30, 2004

1201 New York Ave, N.W., Suite 700
Washington, D.C. 20005
Telephone: (202) 434-1500
Facsimile: (202) 434-1501

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日 2003年10月14日
Date of Application:

出願番号 特願2003-353928
Application Number:

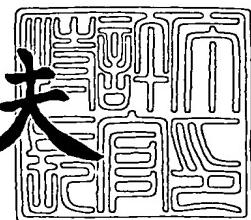
[ST. 10/C] : [JP2003-353928]

出願人 富士通株式会社
Applicant(s):

2004年 1月 7日

特許庁長官
Commissioner,
Japan Patent Office

今井康夫



【書類名】 特許願
【整理番号】 0352285
【提出日】 平成15年10月14日
【あて先】 特許庁長官殿
【国際特許分類】 G06F 17/27
【発明者】
【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社
内
【氏名】 田中 一成
【発明者】
【住所又は居所】 神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社
内
【氏名】 渡部 勇
【特許出願人】
【識別番号】 000005223
【氏名又は名称】 富士通株式会社
【代理人】
【識別番号】 100092152
【弁理士】
【氏名又は名称】 服部 毅巖
【電話番号】 0426-45-6644
【手数料の表示】
【予納台帳番号】 009874
【納付金額】 21,000円
【提出物件の目録】
【物件名】 特許請求の範囲 1
【物件名】 明細書 1
【物件名】 図面 1
【物件名】 要約書 1
【包括委任状番号】 9705176

【書類名】特許請求の範囲**【請求項1】**

複数の文書間の関係を示す図を作成するための関係図作成プログラムにおいて、コンピュータに、

前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出し、

抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算し、

前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置すると共に、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成し、

前記オブジェクトおよび前記関係線で構成される関係図を表示する、
処理を実行させることを特徴とする関係図作成プログラム。

【請求項2】

前記複数の文書の前記オブジェクトを配置する際には、前記時間軸は所定の期間を基準単位とし、異なる前記期間に属する前記文書の前記オブジェクト同士での前記時間軸上の順序を保つことを特徴とする請求項1記載の関係図作成プログラム。

【請求項3】

前記複数の文書として特許文献が入力された場合、

前記特徴素を抽出する際には、前記時間情報として出願日と優先日とを抽出し、

前記複数の文書の前記オブジェクトを配置する際には、前記文書から前記出願日と前記優先日が抽出された場合、前記優先日を前記文書の前記時間情報とみなすことを特徴とする請求項1記載の関係図作成プログラム。

【請求項4】

複数の文書間の関係を示す図を作成するための関係図作成方法において、

前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出し、

抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算し、

前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置すると共に、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成し、

前記オブジェクトおよび前記関係線で構成される関係図を表示する、
ことを特徴とする関係図作成方法。

【請求項5】

複数の文書間の関係を示す図を作成するための関係図作成装置において、

前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出する特徴素抽出手段と、

抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算する関連度計算手段と、

前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置する配置手段と、

計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成する関係線生成手段と、

前記オブジェクトおよび前記関係線で構成される関係図を表示する表示手段と、
を有することを特徴とする関係図作成装置。

【書類名】明細書

【発明の名称】関係図作成プログラム、関係図作成方法、および関係図作成装置

【技術分野】**【0001】**

本発明は関係図作成プログラム、関係図作成方法、および関係図作成装置に関し、特に引用関係や参照関係がない場合でも内容的に関連のある文書の関連付けを示すことができる関係図作成プログラム、関係図作成方法、および関係図作成装置に関する。

【背景技術】**【0002】**

近年、データの記憶媒体の大容量化・低価格化が急速に進んでいる。また、インターネット・インターネットの急速な普及などとともに、世界各地のサーバに格納されている文書をネットワークを介して閲覧することができる。これらの大量の文書情報は、クライアント等の計算機を用いて簡単に収集・蓄積することが可能になってきている。

【0003】

このように集められた情報から必要とする情報を探し出したり、あるいは何らかの知見を得たりするためには、インターネット上の情報量があまりにも膨大である。そのため、利用者の要求に合わせて文書情報を検索・分析することができる検索・分析装置の存在が不可欠となる。

【0004】

文書情報の検索・分析装置としては、利用者によって指定された単語や文字列を含む文書を選択・表示する方法のほか、引用関係や参照関係にある文書を選択・表示する方法、あるいは、時間順序により並べて表示する技術が知られている。たとえば、特許公報等の文献を、公開年毎に並べて表示することができる（たとえば、特許文献1参照）。

【特許文献1】特開2001-92851号公報（第27図）

【発明の開示】**【発明が解決しようとする課題】****【0005】**

また、引用関係や参照関係にある文書を関連付けて関係図を作成する技術も考えられている。たとえば、本出願人は、特願2002-179896号や特願2002-343744号では、文書間の関連付けを線で示すグラフを表示する技術を出願している。引用関係が存在すれば、引用している文書が引用された文書よりも後に書かれていることが明白であるため、文書間の時間的な前後関係を容易に把握できる。

【0006】

しかし、引用関係や参照関係により関係図を作成する技術においては、引用関係や参照関係がなければ関係図を作成できない。そのため、関連が強い文書であっても引用関係や参照関係として記述されていなければ関連性を図示することができない。

【0007】

一方、文書検索の分野においては、文書からキーワードや属性情報などを抽出し、文書間の関連度を計算することもできる。ところが、その場合、文書間の時間的な前後関係が不明確になってしまいうといふ問題点がある。

【0008】

本発明はこのような点に鑑みてなされたものであり、引用関係や参照関係が明示されていない文書間の関連度を明確にした上で、各文書を時系列で表示することができる情報表示プログラムを提供することを目的とする。

【課題を解決するための手段】**【0009】**

本発明では上記課題を解決するために、図1に示すような機能を実現するための関係図作成プログラムが提供される。本発明に係る関係図作成プログラムは、複数の文書1a, 1b, 1c, ... 間の関係を示す図を作成するためのものである。この関係図作成プログラムは、コンピュータに、以下の処理を実行させることができる。

【0010】

コンピュータは、複数の文書1a, 1b, 1c, …それぞれの内容を解析し、時間情報を含む特徴素を抽出する。次に、コンピュータは、抽出された特徴素に基づいて、複数の文書1a, 1b, 1c, …から抽出される文書ペア間の関連度を計算する。さらに、コンピュータは、複数の文書1a, 1b, 1c, …それを示すオブジェクト7a～7gを、時間情報に基づいて時間軸上に配置すると共に、計算された関連度に応じて、文書ペアのオブジェクト7a～7g間を結ぶ関係線を生成する。そして、コンピュータは、オブジェクトおよび関係線で構成される関係図7を表示する。

【0011】

このような関係図作成プログラムをコンピュータに実行させることで、複数の文書から1a, 1b, 1c, …時間情報を含む特徴素が抽出される。すると、複数の文書1a, 1b, 1c, …から抽出された特徴素に基づいて、文書間の関連度が計算される。さらに、複数の文書1a, 1b, 1c, …から抽出された時間情報に基づいて、各文書のオブジェクト7a～7gが時間軸上に配置される。また、文書1a, 1b, 1c, …間の関連度に基づいてオブジェクト間を結ぶ関係線が生成される。そして、各文書のオブジェクト7a～7gと関係線とで構成される関係図7が表示される。

【0012】

また、本発明では上記課題を解決するために、複数の文書間の関係を示す図を作成するための関係図作成方法において、前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出し、抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算し、前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置すると共に、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成し、前記オブジェクトおよび前記関係線で構成される関係図を表示する、ことを特徴とする関係図作成方法が提供される。

【0013】

このような関係図作成方法によれば、複数の文書から時間情報を含む特徴素が抽出される。すると、複数の文書から抽出された特徴素に基づいて、文書間の関連度が計算される。さらに、複数の文書から抽出された時間情報に基づいて、各文書のオブジェクトが時間軸上に配置される。また、文書間の関連度に基づいてオブジェクト間を結ぶ関係線が生成される。そして、各文書のオブジェクトと関係線とで構成される関係図が表示される。

【0014】

さらに、本発明では上記課題を解決するために、複数の文書間の関係を示す図を作成するための関係図作成装置において、前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出する特徴素抽出手段と、抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算する関連度計算手段と、前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置する配置手段と、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成する関係線生成手段と、前記オブジェクトおよび前記関係線で構成される関係図を表示する表示手段と、を有することを特徴とする関係図作成装置が提供される。

【0015】

このような関係図作成装置によれば、特徴素抽出手段により、複数の文書から時間情報を含む特徴素が抽出される。すると、関連度計算手段により文書間の関連度が計算される。さらに、複数の文書から抽出された時間情報に基づいて、各文書のオブジェクトが時間軸上に配置される。また、複数の文書間の関連度に基づいてオブジェクト間を結ぶ関係線が生成される。そして、各文書のオブジェクトと関係線とで構成される関係図が表示される。

【発明の効果】**【0016】**

以上説明したように本発明では、文書から時間情報を含む特徴素を抽出し、特徴素により文書間の関連度を計算し、時間情報に基づいて文書のオブジェクトを時間軸上に配置す

るようしたため、文書間の関係を時系列に沿って容易に把握することができる。

【発明を実施するための最良の形態】

【0017】

以下、本発明の実施の形態を図面を参照して説明する。

本発明は、上記した従来技術の問題点を考慮してなされたものであって、文書間の関連度を計算する技術を利用することにより、引用関係や参照関係がない場合でも関連のある文書を関連付け、時間情報をを利用して関係図を作成できるようにするものである。

【0018】

まず、実施の形態に適用される発明の概要について説明し、その後、実施の形態の具体的な内容を説明する。

図1は、実施の形態に適用される発明の概念図である。本発明に係る関係図作成装置は、複数の文書1a, 1b, 1c, …間の関係を示す図を作成するためのものである。この関係図作成装置は、特徴素抽出手段2、関連度計算手段3、配置手段4、関係線生成手段5、および表示手段6を有している。

【0019】

特徴素抽出手段2は、複数の文書1a, 1b, 1c, …それぞれの内容を解析し、時間情報を含む特徴素を抽出する。特徴素は、たとえば、キーワードや書誌情報である。

関連度計算手段3は、抽出された特徴素に基づいて、複数の文書1a, 1b, 1c, …から抽出される文書ペア間の関連度を計算する。関連度の計算は、たとえば、同じキーワードを多く含む文書同士の関連度を高くする。

【0020】

配置手段4は、複数の文書1a, 1b, 1c, …それぞれを示すオブジェクト7a～7gを、時間情報に基づいて時間軸上に配置する。この際、必ずしも全ての文書間の関係における時間上の前後関係を保つ必要はない。たとえば、少なくとも、関連度がある程度以上の文書ペアにおける文書の時間的な前後関係が保たれるようにすることができる。

【0021】

関係線生成手段5は、計算された関連度に応じて、文書ペアのオブジェクト7a～7g間を結ぶ関係線を生成する。なお、全ての文書ペアのオブジェクト間を関係線で結ばなくてもよい。たとえば、所定の条件に従って文書間の関連付けの間引きを行い、残された関連付けに応じた関係線のみを生成することができる。また、関係線は、関連度に応じて表示形態（色、線の太さなど）を変えることもできる。たとえば、関連度が高い関係を示す関係線を強調表示することができる。

【0022】

表示手段6は、オブジェクトおよび関係線で構成される関係図7を表示する。

このような関係図作成装置によれば、複数の文書1a, 1b, 1c, …が入力されると、特徴素抽出手段2により、複数の文書1a, 1b, 1c, …から時間情報を含む特徴素が抽出される。関連度計算手段3により、複数の文書1a, 1b, 1c, …から抽出された特徴素に基づいて、文書間の関連度が計算される。さらに、配置手段4により、複数の文書1a, 1b, 1c, …から抽出された時間情報に基づいて、各文書のオブジェクト7a～7gが時間軸上に配置される。また、関係線生成手段5により、文書1a, 1b, 1c, …間の関連度に基づいてオブジェクト間を結ぶ関係線が生成される。そして、表示手段6により、各文書のオブジェクト7a～7gと関係線とで構成される関係図7が表示される。

【0023】

たとえば、生成された関係図7は、文書を示す複数のオブジェクト7a～7gが時間軸上に配置されている。オブジェクト7a～7g間は、所定の条件を満たした（間引きされていない）関係線で接続されている。そして、1つの関係線で接続されたオブジェクト同士では、時間情報に従った位置関係で時間軸上に配置されている。たとえば、時間情報で示される期日が後であるほど、右寄りに配置される。

【0024】

このように、文書間の関連度を計算する技術を利用することにより、引用関係や参照関係がない場合でも関連のある文書を関連付け、時間情報を利用して関係図を作成できる。しかも、関連のある文書間の時間的な前後関係を容易に理解することができる。

【0025】

以下、本発明の実施の形態について具体的に説明する。本発明の実施の形態では、ネットワーク経由で大量の文書を収集し、集めた文書に基づいて時系列の関係図を作成するものとする。

【0026】

図2は、ネットワークを介した文書検索を行うシステムの構成例を示す図である。クライアント100は、ネットワーク10を介してサーバ200に接続されている。サーバ200はデータベース210を有している。データベース210には、膨大な量の文書が格納されている。たとえば、特許文献が格納される。

【0027】

クライアント100を使用するユーザは、クライアント100を用いてサーバ200にアクセスし、文書を取得することができる。たとえば、サーバ200に実装されている検索エンジン（データベース検索を実行する機能）に対して、クライアント100から検索要求を送信する。データベースが特許文献であれば、技術用語や国際特許分類のコードなどを検索キーとすることができます。サーバ200は、検索要求に応じて、データベース210の検索を行い、検索条件に合致した文書を検索結果としてクライアント100に返す。

【0028】

クライアント100では、検索結果に含まれる文書を解析し、各文書を示す情報が時系列で並べられた関係図を作成することができる。または、サーバ200において検索結果に含まれる文書を解析し、関係図を作成し、クライアントに送信するようにしてもよい。

【0029】

なお、図2には、1台のサーバ200のみを示しているが、たとえば、インターネット等の広域ネットワークを介して文書検索を行うと、膨大な数のサーバを対象として文書検索を行うことができる。

【0030】

図3は、本発明の実施の形態に用いるクライアントのハードウェア構成例を示す図である。クライアント100は、CPU(Central Processing Unit)101によって装置全体が制御されている。CPU101には、バス107を介してRAM(Random Access Memory)102、ハードディスクドライブ(HDD:Hard Disk Drive)103、グラフィック処理装置104、入力インタフェース105、および通信インターフェース106が接続されている。

【0031】

RAM102には、CPU101に実行させるOS(Operating System)のプログラムやアプリケーションプログラムの少なくとも一部が一時的に格納される。また、RAM102には、CPU101による処理に必要な各種データが格納される。HDD103には、OSやアプリケーションプログラムが格納される。

【0032】

グラフィック処理装置104には、モニタ11が接続されている。グラフィック処理装置104は、CPU101からの命令に従って、画像をモニタ11の画面に表示させる。入力インタフェース105には、キーボード12とマウス13とが接続されている。入力インタフェース105は、キーボード12やマウス13から送られてくる信号を、バス107を介してCPU101に送信する。

【0033】

通信インターフェース106は、ネットワーク10に接続されている。通信インターフェース106は、ネットワーク10を介して、他のコンピュータとの間でデータの送受信を行う。

【0034】

以上のようなハードウェア構成によって、本実施の形態の処理機能を実現することができる。

図4は、関係図作成装置として機能するためのクライアントの機能ブロック図である。クライアント100は、特徴素抽出部110、文書間関連度計算部120、関連付け間引き部130、文書配置計算部140、関係線表示処理部150、および出力処理部160を有している。このクライアント100は、文書情報30が入力されると、関係図作成処理を開始する。

【0035】

特徴素抽出部110は、文書情報30が入力されると、文書情報30から特徴素として、キーワードや書誌情報、時間情報を抽出する。文書毎に抽出された特徴素は、文書間関連度計算部120に渡される。

【0036】

なお、文書情報30は、たとえば、複数の文書の集合である。キーワードは、たとえば、各文書に対して形態素解析を施すことによって抽出することができる。また、書誌情報は、たとえば、文書の作成者等の情報である。なお、文書が特許文献（特許公開公報や特許掲載公報等）の場合、書誌情報として、発明者、出願人、代理人等の情報を抽出することができる。

【0037】

文書から抽出される時間情報は、たとえば、その文書の作成日時や最終更新日時などである。また、文書が特許文献であれば、時間情報として、公開日、登録日、優先権主張日（優先日）等の情報を抽出することができる。

【0038】

文書間関連度計算部120は、抽出された特徴素を利用して文書間の関連度を計算する。具体的には、抽出された特徴素の傾向が似ている文書同士ほど関連度を高くする。たとえば、文書毎に、抽出された特徴素に基づいて特徴を表すベクトルを計算する。そして、各文書のベクトルの近さ（たとえば、ベクトルの内積の大きさ）に応じて、文書間の関連度を計算する。

【0039】

関連付け間引き部130は、文書間関連度計算部120によって関連付けられた中から必要な関連付けだけを選択する。換言すると、関連付けを示す情報のうち、不要は関連付けを破棄する。たとえば、関連度の閾値を設定し、その閾値以上の関連付けのみを選択する。

【0040】

文書配置計算部140は、文書間の関連度を利用して関係図上での文書の配置を決める。具体的には、文書間の関連付けを参照し、関連する文書同士が時系列の前後関係を崩さずに配置を決定する。

【0041】

関係線表示処理部150は、文書間の関連度を利用して関係図上での関係線の表示属性を決定する。たとえば、関係度の高い文書間の関係線を強調するような表示を行う。

出力処理部160は、文書配置計算部140と関係線表示処理部150によって決定された文書配置と関係線表示属性を利用して実際に関係図を表示する。

【0042】

このような構成のクライアント100に対して文書情報30が入力されたときの動作を、以下に示す。

図5は、関係図作成処理の手順を示すフローチャートである。以下、図5に示す処理ステップ番号に沿って説明する。

【0043】

【ステップS11】特徴素抽出部110が、複数の文書31, 32, 33, …を読み込む。

[ステップS12] 特徴素抽出部110は、各文書からキーワード、書誌情報、時間情報等の特徴素を抽出し、特徴素管理テーブル41を作成する。特徴素管理テーブル41には、文書毎に抽出した特徴素の情報が登録される。

【0044】

[ステップS13] 文書間関連度計算部120は、特徴素管理テーブル41を参照し、文書間の関連度計算を行う。関連度計算により関連すると判断された文書間には、文書間関連度情報42が定義される。

【0045】

[ステップS14] 関連付け間引き部130は、文書間関連度情報42を参照し、不要な関連付け情報の間引きを行う。間引きされた関連付け情報は、関連付け間引き情報43に設定される。

【0046】

[ステップS15] 文書配置計算部140は、特徴素管理テーブル41と関連付け間引き情報43とを参照し、時系列に沿って、文書を示すオブジェクトの配置を決定する。

[ステップS16] 関係線表示処理部150は、表示する関連線や線の太さや色といった関連線の表示属性を決定する。

【0047】

[ステップS17] 出力処理部160は、文書配置計算部140によって決定された位置に、各文書を示すオブジェクトを配置し、各オブジェクト間を関係線表示処理部150で決定された表示属性の線で結ぶことで、関係図を生成する。そして、出力処理部160は、生成した関係図をモニタ11に表示する。

【0048】

このようにして関係図において文書を示すオブジェクトを時系列で表示することができる。ところで、時系列で表示することが求められる文書情報として特許文献がある。特許文献は、特許出願の新規性を判断する際の公知技術として利用される場合が多く、そのために公開された日時が非常に重要となる。そのため、特許文献のデータベースから複数の特許文献を抽出した場合、それらを時系列で表示できることが望まれる。そこで、複数の文書31, 32, 33, …として、特許文献を入力した場合を例に採り、図5に示す各ステップの処理の詳細を具体的に説明する。

【0049】

[文書読み込み(ステップS11)]

まず、文書の読み込み処理について具体的に説明する。読み込み対象となる文書には、特許文献のように書誌的事項が含まれている文書がある。

【0050】

図6は、特許文献の例を示している。この図には、入力される特許文献50のフロントページを示している。特許文献50のフロントページには、様々な書誌的事項が記載されている。書誌的事項には、時間情報も含まれる。たとえば、公開日51、出願日52、優先日53がある。

【0051】

クライアント100には、このような特許文献50が複数入力される。たとえば、データベース210から検索した結果として複数の特許文献を取得する。取得した特許文献が、特徴素抽出部110に渡される。

【0052】

[特徴素の抽出(ステップS12)]

特徴素抽出部110において、キーワードおよび書誌情報の抽出が行われる。文書からキーワードを抽出する方法については、様々な技術がある。たとえば、特徴素抽出部110は、文書内の文章を語句毎に分割する。次に、特徴素抽出部110は、各語句の品詞を判断する。そして、特徴素抽出部110は、特定の品詞(たとえば、名詞や動詞等)のキーワードとして抽出する。どのような品詞を抽出するのかは、ユーザが任意に設定することができる。たとえば、特徴素抽出部110は、モニタ11に品詞設定画面を表示し、ユ

ーザは、その画面上で特徴素として抽出する品詞を指定することができる。

【0053】

図7は、品詞設定画面の一例を示す図である。品詞設定画面60には、直前の設定ボタン62、デフォルトボタン63、クリアボタン64、全選択ボタン65、設定ボタン66、およびキャンセルボタン67が設けられている。

【0054】

品詞選択部61は、文書に対して形態素解析を行うことで取得される語句の品詞のリストが表示される。この例では、書誌情報の項目名も品詞の1つとして扱われ、品詞選択部61に表示されている。ユーザは、品詞選択部61から、キーワードとして抽出すべき品詞を選択することができる。

【0055】

直前の設定ボタン62は、抽出すべきキーワードの品詞の設定内容を変更後、設定内容を直前の設定に戻すためのボタンである。誤った設定操作を行った場合、直前の設定ボタン62を押すことで、操作入力前の状態に設定内容を戻すことができる。

【0056】

デフォルトボタン63は、抽出すべきキーワードの品詞として、予め指定されている品詞を設定するためのボタンである。クライアント100には、抽出対象品詞の初期値が予め設定されており、デフォルトボタン63が押されると、初期値として設定されている品詞のみが、抽出すべきキーワードの品詞として設定される。

【0057】

クリアボタン64は、品詞選択部61で選択されている品詞を、非選択の状態に変更するためのボタンである。

全選択ボタン65は、品詞選択部61内の全ての品詞を選択するためのボタンである。

【0058】

設定ボタン66は、品詞選択部61で選択されている品詞を、抽出対象とする品詞として設定するためのボタンである。

キャンセルボタン67は、抽出対象とする品詞の設定を変更せずに、品詞設定画面60を閉じるためのボタンである。

【0059】

なお、品詞選択部61は、全ての品詞が一画面で表示できないとき、スクロールバーを用いて品詞選択部61の表示内容をスクロールさせることで、各品詞を表示させることができる。

【0060】

図8は、品詞選択部スクロール後の品詞設定画面の例を示す図である。図8に示すように、品詞選択部61の表示内容は、スクロールさせることができる。

このように、品詞設定画面60により、文書間の関連度計算に使用するための任意の品詞を指定することができる。たとえば、デフォルトでは、名詞や固有名詞などのキーワードになりうる品詞にしておき、IPC(国際特許分類)や出願人などを任意に選択することでこれらの情報をを利用して文書の関連度を計算することができる。

【0061】

なお、文書から書誌情報を抽出する方法については、様々な方法が考えられる。たとえば、文書ファイルには、プロファイルとして作成者や作成日時が登録されている場合がある。これらのプロファイルの内容を書誌情報として抽出することもできる。

【0062】

特許文献のように、文書内に書誌情報の項目が設けられている場合、各項目の種別(発明者、出願人等)を判別して、その項目に登録されている情報を書誌情報として抽出することができる。書誌情報には、時間情報も含まれる。特許文献の時間情報としては、出願日、優先日、公開日、登録日などがある。

【0063】

特徴素抽出部110は、抽出したキーワードや書誌情報により特徴素管理テーブル41

を作成する。

図9は、特徴素管理テーブルのデータ構造例を示す図である。特徴素管理テーブル41には、キーワード、書誌情報、および時間情報に分類して、各文書の特徴素が登録されている。

【0064】

たとえば、キーワードの分類項目には、キーワードを示す文字列と品詞（名詞、動詞など）との組が登録されている。書誌情報の分類項目には、書誌的事項の項目名とその内容との組が登録される。文書が特許文献であれば、書誌情報として、発明者、出願人などが登録される。時間情報の分類項目には、時間に関する項目と、その項目に設定された日付あるいは日時が登録される。文書が特許文献であれば、時間情報として、出願日、優先日、公開日などが登録される。

ここでは、抽出する特徴素の品詞を選択するとして説明したが、特徴素抽出部110では全ての品詞を抽出しておいて、文書間関連度計算部120において文書間の関連度を計算する際に選択された品詞の特徴素のみを使って計算を行うようにしてもよい。

【0065】

【文書関連度の計算（ステップS13）】

その後、文書毎の特徴素管理テーブル41を利用して、文書間関連度が計算される。たとえば、特徴素管理テーブル41内のキーワードから文書一単語行列を定義する。

【0066】

図10は、文書一単語行列の例を示す図である。文書一単語行列41aでは、列（縦）方向に文書名が設定され、行（横）方向にキーワードが設定されている。文書名とキーワードとの交差する領域には、その文書内に含まれるキーワードの数が設定される。

【0067】

なお、図10の例では、最も単純な値としてそれぞれの文書での単語の数を入れているが、それぞれの文書でのキーワードの重みなどをいれてもよい。また、関連度計算において使用する品詞を区別できるように、それぞれのキーワードには品詞名がつけられている。

【0068】

このような、文書一単語行列を用いて、文書間の関連度を計算することができる。文書間の関連度の計算方法については既知の技術で実現可能である。たとえば、ベクトル空間法(vector-space model)と呼ばれる関連度の計算方法がある。ベクトル空間法は、各文書の特徴をある統一的な表現によって表し、この間に類似度(similarity)を定義することで、似た文書を探し出す方法である。

【0069】

すなわち、各文書の特徴をベクトルで表現する。ベクトルは、その文書から抽出された特徴素に応じて決定される。2つの文書間の類似度は、それぞれの文書に対応するベクトルの内積で求めることができる。ベクトルの内積の値が大きいほど、類似していることを示す。ベクトル空間法における類似度を文書間の関連度と見なすことで、文書間の関連度が求められる。ベクトル空間法の詳細は、“長尾真、佐藤理史、黒橋禎夫、角田達彦、「自然言語処理」、岩波書店、1996年4月26日、P421-P424”に記載されている。

【0070】

ところで、文書間の関連度を計算するときに、どの特徴素を利用するのかが問題となる。一般的には文書からキーワードを抽出し、そのキーワードを特徴素として使用する方法が良く知られている。ただし、本実施の形態における文書間の関連度を計算する方法としては、キーワードのみに限らず書誌情報を使用することも考えられる。

【0071】

ここで言う書誌情報とは、たとえば、特許文献を対象にするような場合には出願人や発明者、IPCなどの分類コードなどである。また、特許文献以外であっても文書の作者名や、文書の作者の所属など文書につけられている情報を利用することができる。社内分類のように別途追加された情報を利用してもよい。

【0072】

更に、文書から抽出する情報としてもキーワードばかりではなく、係り受け組などのよ
うなキーワード間の関連までを含んだものや、特定の規則にしたがって抽出した特徴情報
など文書を特徴付ける様々な情報を活用することができる。

【0073】

なお、ここでは関連性の度合いを示す指標として関連度という言葉を使っているが、必
ずしも連続的な値ではなく、複数の文書で共有する特徴情報があるかないかといった0、
1の関係でもよい。

【0074】

算出された文書間の関連度は、文書間関連度情報42に登録される。

図11は、文書間関連度情報のデータ構造例を示す図である。文書間関連度情報42で
は、たとえば、文書ペアと関連度との項目が設けられる。

【0075】

文書ペアの項目には、関連度の比較対象となる2つの文書が登録される。この項目には
、入力された文書31, 32, 33, …から2つの文書を選択したときの全ての組み
合わせが登録される。関連度の項目には、対応する比較文書間の関連度が登録される。

【0076】**[文書関係の間引き（ステップS14）]**

次に、文書間の関連度情報から関係図の作成に必要な関連だけを選別（間引き）する方
法について説明する。

【0077】

文書配置計算や関係線の表示属性に活用する文書の関連を選択する場合には、関係図上
の文書がばらばらになってしまふと文書間の関係がわからない。そのため、すべての文書
が最低1つは別の文書と接続されるように関連付けを選択する必要がある。

【0078】

たとえば、以下の方法により文書間の関連度情報から関連付けを選択する。

図12は、文書関係間引き処理の手順を示すフローチャートである。以下、図12に示
す処理をステップ番号に沿って説明する。

【0079】

[ステップS21] 関連付け間引き部130は、全ての文書31, 32, 33, …
のペアについて関連度順にソートする。ソートされた文書ペアには、関連度が高い順に、
0から始まる昇順の番号が振られる。また、初期状態では、1文書を1グループと考える
。

【0080】

[ステップS22] 関連付け間引き部130は、変数iに0を設定する。

[ステップS23] 関連付け間引き部130は、i番目の文書ペアの各文書が別グル
ープに属しているか否かを判断する。別のグループに属している場合、処理がステップS2
4に進められる。同一のグループに属している場合、処理がステップS25に進められる
。

【0081】

[ステップS24] 関連付け間引き部130は、i番目の文書ペア間の関連を有効にす
る。具体的には、関連付け間引き部130は、関連付け間引き情報43におけるi番目の
文書ペアに対応する欄に、関連付けが有効であることを示す情報を設定する。関連付けが
有効になった文書それぞれのグループは統合され、1つのグループとなる。

【0082】

[ステップS25] 関連付け間引き部130は、全ての文書が同一グループに属してい
るか否かを判断する。全ての文書が同一グループに属していれば、処理がステップS27
に進められる。複数のグループが存在していれば、処理がステップS26に進められる。

【0083】

[ステップS26] 関連付け間引き部130は、iの値をインクリメント（1だけ加算

) する。その後、処理がステップS23に進められ、次の順番の文書ペアについて、関連の有効性を判断する。これにより、全ての文書が同一グループになるまで、ステップS23～ステップS25の処理が繰り返される。

【0084】

このように、ステップS21～ステップS26の処理により、全ての文書のうち同じグループでない文書ペアの中で最も関連の強い文書ペア間の関連が有効とされ、同じグループとなる。その後、同じグループに属した文書同士の文書ペアが検討されたとき、すでに同じグループに属していることから、文書ペアの関連は有効とされない。

【0085】

[ステップS27] 関連付け間引き部130は、変数iに0を設定する。

[ステップS28] 関連付け間引き部130は、i番目の文書ペア間の関係を有効にする。

【0086】

[ステップS29] 関連付け間引き部130は、関連が有効である文書ペアの数が所定値に達しているか否かを判断する。関連が有効である文書ペアの数が所定値に達していれば、処理が終了する。所定値に達していなければ、処理がステップS30に進められる。

【0087】

[ステップS30] 関連付け間引き部130は、iの値をインクリメント(1だけ加算)する。その後、処理がステップS28に進められる。

このように、ステップS28～ステップS30では、ステップS23, S24で示される「同じグループ内では関連が強くても関連を有効としない」という条件をなくして、関連度が高い順に有効な関連が一定数(たとえば全ての関連の数十%)になるまで関連を有効にすることができます。その結果、関連付け間引き情報43が生成される。

【0088】

図13は、関連付け間引き情報のデータ構造例を示す図である。関連付け間引き情報43は、関連度によるソート後の文書間関連度情報42aの各文書ペアに対応する選択情報を有している。選択情報には、関連が有効として選択された文書ペアに対して「有効」が設定されている。すなわち、「有効」と設定された文書ペア間の関連付けが選択され、なにも設定されていない文書ペア間の関連付けは間引きされる。

【0089】

ところで、どの程度の関連度以下の関連付けを間引きするのか(間引きする関連度の閾値)は、ユーザが任意に設定することができる。

図14は、間引きの設定画面の例を示す図である。図14の示す間引き設定画面90の例では、間引き条件として、「エッジの数」「関連度」「平均次数」といった値によって制御することが可能である。それぞれの条件については以下のよう意味を持つ。

【0090】

「エッジの数」は、全ての関係線数の何%の関係線を残すかを示す。チェックボックス91aにチェックマークが表示されているときに、「エッジの数」の条件が有効となる。残すべき関係線の割合は、テキストボックス91bにパーセンテージによって入力できる。

【0091】

「関連度」は、文書間の関連度の値がいくつ以下のものを間引くかを示す。チェックボックス92aにチェックマークが表示されているときに、「関連度」の条件が有効となる。関連度の閾値となる値は、テキストボックス92bに数値によって入力できる。

【0092】

「平均次数」は、1つの文書あたり平均いくつの関係線が残るようにするかを示す。チェックボックス93aにチェックマークが表示されているときに、「平均次数」の条件が有効となる。平均次数は、テキストボックス93bに数値によって入力できる。

【0093】

また、図14の例では、チェックボックス94により連結性を保持するかどうかの選択

も行える。連結性を保持するようにすれば、間引き後もすべての文書は最低1つは別の文書とつながれることになる。

【0094】

さらに、チェックボックス95により、間引きしたエッジ（関係線）を透明化するかどうかの選択も可能である。間引きした関係線を透明化すれば、関連度の高い関係線のみが表示され、文書間の関連状況が把握しやすくなる。

【0095】

「1ノード当たりの最大エッジ数」を設定するためのテキストボックス96も設けられている。このテキストボックス96は、1つの文書に線が集中して放射状の図ができるのを防ぐためのものである。1つの文書に線が集中するような場合であっても、このテキストボックス96で指定した値以上に線がつながれないように制限される。

【0096】

これらの制限の基に間引きを行う方法については、本出願人により特許出願されている（特願2002-179896号）。

間引き設定内容は、デフォルト値（初期値）を予め設定しておくことができる。間引き設定画面90が最初に表示されたときは、デフォルト値が設定された状態で画面表示される。図14には、デフォルト値が示されているものとする。この例では、間引き条件として「平均次数」が選択され、その平均次数に「3」が設定されている。また、1ノード当たりの最大エッジ数に「5」が設定されている。

【0097】

間引き設定画面90には、OKボタン97とキャンセルボタン98とが設けられている。OKボタン97が押されると、そのとき間引き設定画面90上で設定されている条件が確定する。キャンセルボタン98が押されると、設定内容が変更されずに間引き設定画面が閉じる。

【0098】

このようにして、ユーザは任意の間引き条件を設定することができる。

なお、間引きの方法は、他にも様々な方法が考えられる。たとえば、文書に接続可能な関係線の数（その文書を含む文書ペアのうち関連が有効とされる数）を制限することができる。

【0099】

また、文書ペア間の関係において代替パス（他の文書を経由して関連付けられていること）が存在する場合に、その文書ペア間の直接の関連を無効にすることもできる。このような技術は、本出願人により特許出願されている（特願2002-343744号）。

【0100】

このようにして、関連の間引きを行うことができる。その結果、関係図を表示したときに、重要な関連線のみで文書同士が接続され、ユーザが文書間の関係を理解しやすくなる。

【0101】

[文書配置計算（ステップS15）]

次に、文書配置計算について説明する。ここでは、関連付けられた文書間のみで時間順序が保存されるように文書をレイアウトする方法について説明する。文書配置の詳細な計算法については“杉山公造、「グラフ自動描画法とその応用」、コロナ社、1993年”に記載された方法を用いることができる。以下、前記文献に記載された方法のうち、比較的簡単な例について説明する。

【0102】

図15は、文書配置計算処理の手順を示すフローチャートである。以下、図15に示す処理をステップ番号に沿って説明する。なお、本実施の形態では、関係図の横方向に時間軸をとり、左から右に時間が進むものとする。

【0103】

[ステップS41] 文書配置計算部140は、文書を示すオブジェクト（以下、文書オ

プロジェクトと呼ぶ)をランダムに配置し、特徴素管理テーブル41と関連付け間引き情報43を利用して、文書オブジェクト間の関連付けを示す矢印の作成を行う。すなわち、関連付け間引き情報43において有効になっている関係にある文書ペアで時間的に古い文書の文書オブジェクトから新しい文書の文書オブジェクトを指示示す矢印を作成する。

【0104】

[ステップS42] 文書配置計算部140は、各文書に付けられた矢印が全て右側(時間の進行方向)を向くように全ての文書を直線上に並べる。

[ステップS43] 文書配置計算部140は、文書オブジェクトを並べ終わったら、各文書オブジェクトに階層を割り当てる。具体的には、文書配置計算部140は、「左側に直列につながっている文書オブジェクトの最大階層値+1」をその文書オブジェクトの階層を示す値とする。

【0105】

[ステップS44] 文書配置計算部140は、最後に文書オブジェクトの配置を決定する。すなわち、文書配置計算部140は、文書オブジェクトを配置する空間を階層に分け、文書オブジェクトにつけられた階層によって文書オブジェクトの横方向の配置を決定する。文書オブジェクトの縦方向の位置は、「関連の強い文書が近くになるようにする」、「文書オブジェクト間の関係線の交差が少ないようにする」といった条件に基づいて決定する。

【0106】

以下、図16～図19を参照して、文書オブジェクトの配置例を説明する。なお、図16～図19では、各文書オブジェクトを丸印で表し、丸印の中に各文書オブジェクトの識別番号を示すものとする。

【0107】

図16は、ランダムに配置された文書オブジェクトを示す図である。この例では、12個の文書オブジェクト71～82が配置されている。文書オブジェクト71～82は、それぞれ少なくとも1つの他の文書オブジェクトとの間で関連付けが有効となっている。そして、有効な関連付けが矢印で示されている。矢印を設定する際には、関連する2つの文書の時間情報が比較され、時間が古い文書の文書オブジェクトから新しい文書の文書オブジェクトに向かって矢印が設定される。たとえば、識別番号「1」の文書オブジェクト71と識別番号「5」の文書オブジェクト75との関連付けが有効であり、文書オブジェクト71の方が文書オブジェクト75よりも古い時間情報が設定されている。

【0108】

その後、各文書オブジェクト71～82が、時間軸に沿って並べられる。

図17は、時間軸に沿って並べられた文書オブジェクトを示す図である。各文書オブジェクト71～82が時間軸に沿って並べられることで、全ての矢印は、時間進行に沿った方向を指し示すことになる。

【0109】

その後、各文書オブジェクト71～82に関して、階層が決定される。本実施の形態では、判断対象となる文書オブジェクトの左側に関係付けられている文書オブジェクトの階層を示す値に1を加算した値が、その文書オブジェクトの階層となる。たとえば、識別番号「5」の文書オブジェクト75は、左側に階層「1」の文書オブジェクト71が関連付けられているため、階層は「2」である。また、識別番号「12」の文書オブジェクト82は、左側に2つの文書オブジェクト77, 79が関連付けられている。文書オブジェクト77の階層は「2」であり、文書オブジェクト79の階層は「3」である。このように左側に複数の文書が関連付けられている場合、高い方の階層に1を加算した値が、判断対象の文書オブジェクト82の階層となる。したがって、文書オブジェクト82の階層は「4」である。

【0110】

図18は、階層毎に配置された文書オブジェクトを示す図である。時間軸に沿って領域が区切られており、各領域に階層が割り振られている。このとき、番号が大きい階層ほど

、時間軸上で新しい方の領域に割り振られる。

【0111】

その後、文書オブジェクトの縦方向の位置を決定する。具体的には、文書オブジェクト間の関係線の交差が少ないよう、文書オブジェクトの縦方向の位置が決定される。

図19は、決定された位置に文書オブジェクトが配置された関係図を示す図である。この例では、階層「1」の文書オブジェクトの配置が、上から文書オブジェクト74、文書オブジェクト71、文書オブジェクト73、文書オブジェクト72の順に変更されている。また、階層「3」の文書の配置が、上から文書オブジェクト79、文書オブジェクト78、文書オブジェクト80の順に変更されている。このようにして、文書が時系列で配置される。

【0112】

[関連線表示属性決定（ステップS16）]

次に、文書間の関係線の表示属性に反映させる方法について説明する。

文書間の有効にした関連付けに対し、たとえば、以下の方法で関係線の表示属性に反映させることができる。

【0113】

有効な関連と、その他の関連を、それぞれ別の表示属性（たとえば色や太さ）によって表現する。たとえば、有効な関連付けを表す関係線を強調表示する。強調表示の方法としては、輝度を高くしたり、線を太くしたり、線を赤などの目立つ色で表示したりする方法がある。

【0114】

また、有効な関連以外をすべて非表示とすることもできる。具体的には、ユーザは、間引き設定画面90において、チェックボックス95を選択することで、間引きしたエッジ（関係線）を透明化させることができる。

【0115】

[マップ表示（ステップS17）]

生成された関係図は、出力処理部160によってマップ表示される。

図20は、関係図の表示例を示す図である。これは、特許文献の集合を入力したときの例である。図20には、それが個別の特許文献を表す7つの文書オブジェクトが表示されている。

【0116】

文書オブジェクト201は、文書オブジェクト202, 203, 206との間で有効な関連付けを有している。このとき、文書オブジェクト201は、有効な関連付けが成されている文書オブジェクト202, 203, 206の全てよりも時間的に古い時間情報を有している。

【0117】

文書オブジェクト202は、文書オブジェクト205, 206との間で有効な関連付けを有している。このとき、文書オブジェクト202は、有効な関連付けが成されている文書オブジェクト205, 206のいずれよりも時間的に古い時間情報を有している。また、文書オブジェクト202と文書オブジェクト205との間は、比較的関連度が高いため、太い関係線で接続されている。

【0118】

文書オブジェクト203は、文書オブジェクト205, 206との間で有効な関連付けを有している。このとき、文書オブジェクト203は、有効な関連付けが成されている文書オブジェクト205, 206のいずれよりも時間的に古い時間情報を有している。

【0119】

文書オブジェクト204は、文書オブジェクト205との間で有効な関連付けを有している。このとき、文書オブジェクト204は、有効な関連付けが成されている文書オブジェクト205よりも時間的に古い時間情報を有している。

【0120】

文書オブジェクト205は、文書オブジェクト206, 207との間で有効な関連付けを有している。このとき、文書オブジェクト205は、有効な関連付けが成されている文書オブジェクト206, 207のいずれよりも時間的に古い時間情報を有している。また、文書オブジェクト205と文書オブジェクト206との間は、比較的関連度が高いため、太い関係線で接続されている。

【0121】

文書オブジェクト206は、文書オブジェクト207との間で有効な関連付けを有している。このとき、文書オブジェクト206は、有効な関連付けが成されている文書オブジェクト207よりも時間的に古い時間情報を有している。

【0122】

以上のようにして、文書同士の関係を示す線によって、文書を表すオブジェクト同士を結び、各文書オブジェクトを時系列に表示することができる。

【その他の応用例】

上記の説明では、文書間の有効な関連付け（間引き後の関連情報）に応じて、文書配置を計算しているが、間引き前の文書間関連情報を使用して文書配置計算を行うこともできる。また、関連線表示属性に関しても同様に、間引き前の関連情報を使用して決定することもできるし、間引き後の関連情報を使用して決定することもできる。

【0123】

つまり、文書配置計算と関連線表示属性計算との方法（配置等計算法）については、以下の4通りの方法が考えられる。

（配置等計算法a）間引き前の文書間関連情報で文書配置計算と関連線表示属性の決定とも行う。

【0124】

（配置等計算法b）間引き後の文書間関連情報で文書配置計算と関連線表示属性の決定とも行う。

（配置等計算法c）間引き前の文書間関連情報で文書配置計算を行い、間引き後の文書間関連情報で関連線表示属性の決定を行う。

【0125】

（配置等計算法d）間引き後の文書間関連情報で文書配置計算を行い、間引き前の文書間関連情報で関連線表示属性の決定を行う。

文書配置の具体的な計算法については、「杉山公造、「グラフ自動描画法とその応用」、コロナ社、1993年」に記載された方法を用いることができる。

【0126】

また、関係図では時間に沿った順序に並べる必要がある。時間順序を保存した関係図の作成方法としては以下のようなものが考えられる。

（時間順保存法1）関連付けられた文書間のみで時間順序が保存されるように文書オブジェクトをレイアウトする。

【0127】

（時間順保存法2）関係図全体で時間順序が保存されるように文書オブジェクトをレイアウトする。

（時間順保存法3）年または月または日の単位で時間順序が保存されるように文書オブジェクトをレイアウトする。

【0128】

これら文書配置計算や関連線表示属性の決定方法と、時間順序を保存した関係図の作成方法とは、それぞれ任意に組み合わせることができる。

たとえば、図20に示したのは、（配置等計算法b）と（時間順保存法1）を組み合わせた場合の例である。すなわち、間引き後の関連により配置が計算され、関連線が付けられている文書間において時間順序が保存されるように文書の配置が決定されている。そして、間引きを行った後に残った関連付けによって文書間に関係線が引かれている。

【0129】

また、（配置等計算法b）と（時間順保存法2）とを組み合わせて、関係図を作成することもできる。

図21は、全ての文書間の時間順序を保存した関係図の例を示す図である。図21では、互いの関連付けが間引きされている文書オブジェクト202、文書オブジェクト203、文書オブジェクト204それぞれの間においても、時間順序が保存されている。すなわち、文書オブジェクト201～207は、時間情報として早い期日が設定されている文書ほど、図中の左寄りに表示されている。

【0130】

さらに、（配置等計算法d）と（時間順保存法1）とを組み合わせて関係図を作成することもできる。

図22は、間引き前の関連付けを示す関係線を含めて表示した関係図の例を示す図である。図22では、文書オブジェクト201～207の配置は図20と同じであるが、表示される関連線が間引き前であるため関連線の数が多くなっている。なお、間引き対象となる関係線は、他の関係線よりも細く表示されている。

【0131】

文書の配置計算に使われる時間情報としては、文書の作成日や更新日といった文書の附加情報が考えられる。また、特許文献の出願日や公開日、優先日のように書誌情報として文書に付加されている場合にはこれらの情報を抽出して時間情報として使用してもよい。

【0132】

更に、関係図の配置を決めるときには一種類の時間情報を単独で使うだけではなく、たとえば、優先日がある場合には優先日を優先し、優先日が無い場合には出願日を使用するといった具合に複数の時間情報を組み合わせて使用するということも可能である。このような時間情報の組み合わせは特許文献に限ったものではなく、様々な文書に適用できる。たとえば、打ち合わせ資料や手続き書類の作成日が同じ場合には、更新日を使うことによって資料の更新履歴の関係図や手続きフローの自動作成に応用することも可能である。

【0133】

文書間の引用関係や参照関係があった場合これらを利用して以下のように処理することもできる。なお、文書の引用関係や参照関係は、特徴素の1つとして特徴素抽出部110において抽出することができる。

【0134】

図23は、文書間の引用関係や参照関係を含む場合の文書関係の間引き手順を示すフローチャートである。図23に示す処理は、ステップS51を除き図12に示す処理と同様である。すなわち、ステップS52～ステップS61の各処理は、それぞれ図12のステップS21～ステップS30の処理と同じである。

【0135】

関連付け間引き部130は、最初に、引用関係または参照関係にある文書間の関連付けを有効にする（ステップS51）。これにより、引用関係または参照関係にある文書間の関連付けは間引きの対象から除外される。その後、処理がステップS52に進められ、図12と同様の手順で関連付けの有効化が行われる。

【0136】

また、有効な関連を文書配置計算に反映させる方法として、上述の方法以外に次のような方法がある。たとえば、特に引用関係や参照関係によって有効になった関連を利用して文書配置計算を行うこともできる。

【0137】

また、有効な関連を文書間の関係線の表示属性に反映させる方法として、次のような方法が考えられる。たとえば、特に引用関係や参照関係がある関連だけ別の表示属性（たとえば色や太さ）によって表現することができる。

関係図を表示する時には、文書の書誌情報などを利用して対応する文書オブジェクトの表示属性（たとえば枠の色や背景の色）を変えて表示してもよい。たとえば、特許文献の場合であれば、出願人やIPCが同じ特許文献を表すオブジェクトの枠の色を同じにするな

どしてもよい。

【0138】

なお、上記の処理機能は、コンピュータによって実現することができる。その場合、クライアントが有すべき機能の処理内容を記述したプログラムが提供される。そのプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピュータで読み取り可能な記録媒体としては、磁気記録装置、光ディスク、光磁気記録媒体、半導体メモリなどがある。磁気記録装置には、ハードディスク装置（HDD）、フレキシブルディスク（FD）、磁気テープなどがある。光ディスクには、DVD（Digital Versatile Disc）、DVD-RAM（Random Access Memory）、CD-ROM（Compact Disc Read Only Memory）、CD-R（Recordable）／RW（ReWritable）などがある。光磁気記録媒体には、MO（Magneto-Optical disk）などがある。

【0139】

プログラムを流通させる場合には、たとえば、そのプログラムが記録されたDVD、CD-ROMなどの可搬型記録媒体が販売される。また、プログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することもできる。

【0140】

プログラムを実行するコンピュータは、たとえば、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、自己の記憶装置に格納する。そして、コンピュータは、自己の記憶装置からプログラムを読み取り、プログラムに従った処理を実行する。なお、コンピュータは、可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することもできる。また、コンピュータは、サーバコンピュータからプログラムが転送される毎に、逐次、受け取ったプログラムに従った処理を実行することもできる。

【0141】

（付記1） 複数の文書間の関係を示す図を作成するための関係図作成プログラムにおいて、

コンピュータに、

前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出し、

抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算し、

前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置すると共に、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成し、

前記オブジェクトおよび前記関係線で構成される関係図を表示する、
処理を実行させることを特徴とする関係図作成プログラム。

【0142】

（付記2） 前記関係線を生成する際には、引用関係のない前記文書ペアの前記関連度に基づいて、所定の前記文書ペア間の前記関係線を間引くことを特徴とする付記1記載の関係図作成プログラム。

【0143】

（付記3） 前記関係線を生成する際には、前記引用関係がある前記文書ペアの前記関係線を、他の前記関係線とは異なる表示形態で表示することを特徴とする付記3記載の関係図作成プログラム。

【0144】

（付記4） 前記複数の文書の前記オブジェクトを配置する際には、少なくとも、前記関連性のある前記文書ペアの各前記オブジェクトを、前記時間軸方向に前記時間情報に基づく順番に配置することを特徴とする付記1記載の関係図作成プログラム。

【0145】

(付記5) 前記複数の文書の前記オブジェクトを配置する際には、前記複数の文書それぞれの前記オブジェクトを、前記時間軸方向に前記時間情報に基づく順番で配置することを特徴とする付記1記載の関係図作成プログラム。

【0146】

(付記6) 前記複数の文書の前記オブジェクトを配置する際には、前記時間軸は所定の期間を基準単位とし、異なる前記期間に属する前記文書の前記オブジェクト同士での前記時間軸上の順序を保つことを特徴とする付記1記載の関係図作成プログラム。

【0147】

(付記7) 前記複数の文書として特許文献が入力された場合、前記特徴素を抽出する際には、前記時間情報として出願日を抽出することを特徴とする付記1記載の関係図作成プログラム。

【0148】

(付記8) 前記複数の文書として特許文献が入力された場合、前記特徴素を抽出する際には、前記時間情報として出願日と優先日とを抽出し、前記複数の文書の前記オブジェクトを配置する際には、前記文書から前記出願日と前記優先日が抽出された場合、前記優先日を前記文書の前記時間情報とみなすことを特徴とする付記1記載の関係図作成プログラム。

【0149】

(付記9) 複数の文書間の関係を示す図を作成するための関係図作成方法において、前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出し、抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算し、前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置すると共に、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成し、前記オブジェクトおよび前記関係線で構成される関係図を表示する、ことを特徴とする関係図作成方法。

【0150】

(付記10) 複数の文書間の関係を示す図を作成するための関係図作成装置において、前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出する特徴素抽出手段と、抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算する関連度計算手段と、前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置する配置手段と、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線生成手段と、前記オブジェクトおよび前記関係線で構成される関係図を表示する表示手段と、を有することを特徴とする関係図作成装置。

【0151】

(付記11) 複数の文書間の関係を示す図を作成するための関係図作成プログラムを記録したコンピュータ読み取り可能な記録媒体において、コンピュータに、前記複数の文書それぞれの内容を解析し、時間情報を含む特徴素を抽出し、抽出された前記特徴素に基づいて、前記複数の文書から抽出される文書ペア間の関連度を計算し、前記複数の文書それぞれを示すオブジェクトを、前記時間情報に基づいて時間軸上に配置すると共に、計算された前記関連度に応じて、前記文書ペアの前記オブジェクト間を結ぶ関係線を生成し、

前記オブジェクトおよび前記関係線で構成される関係図を表示する、
処理を実行させることを特徴とする関係図作成プログラムを記録したコンピュータ読み
取り可能な記録媒体。

【図面の簡単な説明】

【0152】

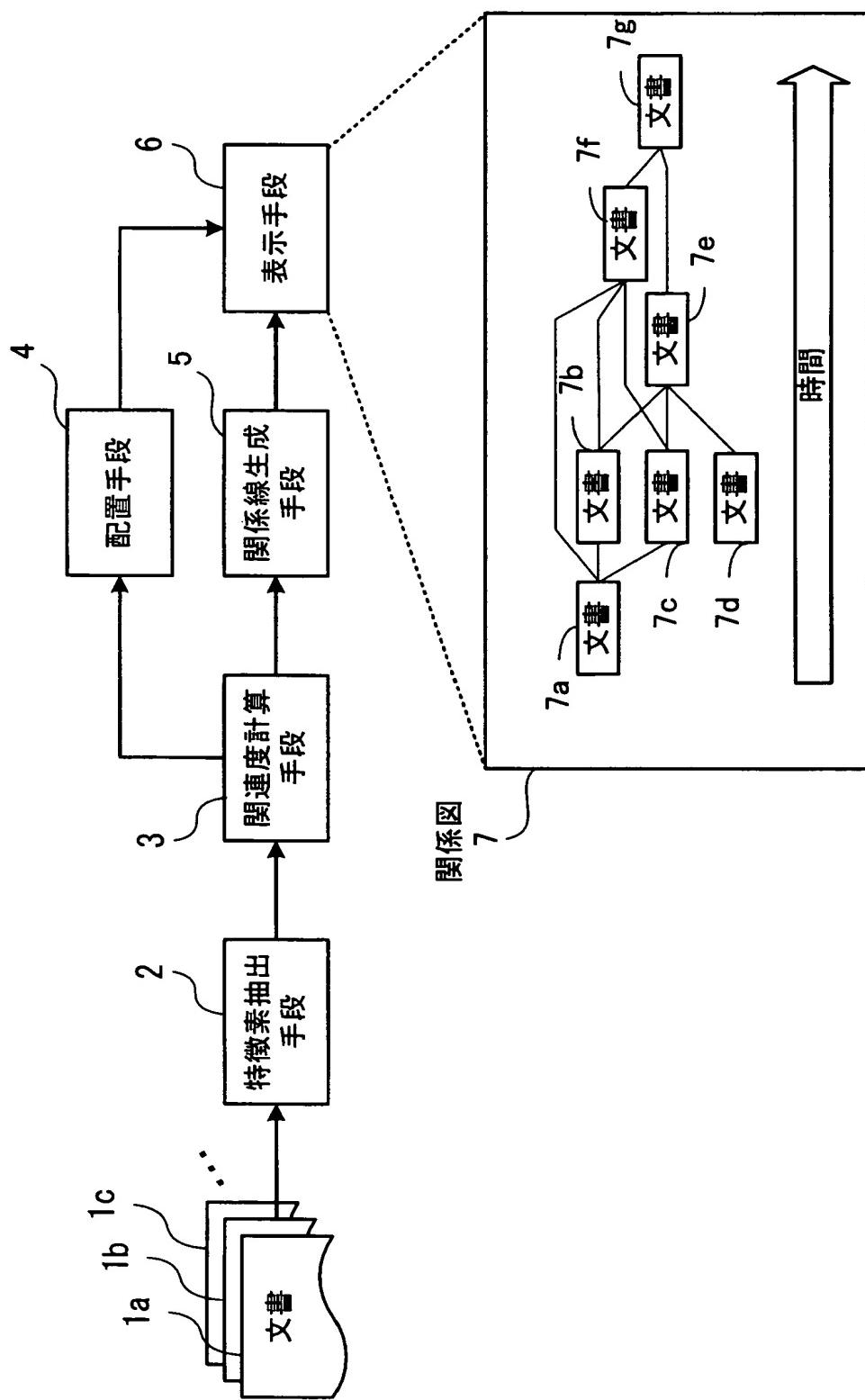
- 【図1】実施の形態に適用される発明の概念図である。
- 【図2】ネットワークを介した文書検索を行うシステムの構成例を示す図である。
- 【図3】本発明の実施の形態に用いるクライアントのハードウェア構成例を示す図である。
- 【図4】関係図作成装置として機能するためのクライアントの機能ブロック図である。
- 【図5】関係図作成処理の手順を示すフローチャートである。
- 【図6】特許文献の例を示している。
- 【図7】品詞設定画面の一例を示す図である。
- 【図8】品詞選択部スクロール後の品詞設定画面の例を示す図である。
- 【図9】特徴素管理テーブルのデータ構造例を示す図である。
- 【図10】文書一単語行列の例を示す図である。
- 【図11】文書間関連度情報のデータ構造例を示す図である。
- 【図12】文書関係間引き処理の手順を示すフローチャートである。
- 【図13】関連付け間引き情報のデータ構造例を示す図である。
- 【図14】間引きの設定画面の例を示す図である。
- 【図15】文書配置計算処理の手順を示すフローチャートである。
- 【図16】ランダムに配置された文書オブジェクトを示す図である。
- 【図17】時間軸に沿って並べられた文書オブジェクトを示す図である。
- 【図18】階層毎に配置された文書オブジェクトを示す図である。
- 【図19】決定された位置に文書オブジェクトが配置された関係図を示す図である。
- 【図20】関係図の表示例を示す図である。
- 【図21】全ての文書間の時間順序を保存した関係図の例を示す図である。
- 【図22】間引き前の関連付けを示す関係線を含めて表示した関係図の例を示す図である。
- 【図23】文書間の引用関係や参照関係を含む場合の文書関係の間引き手順を示すフローチャートである。

【符号の説明】

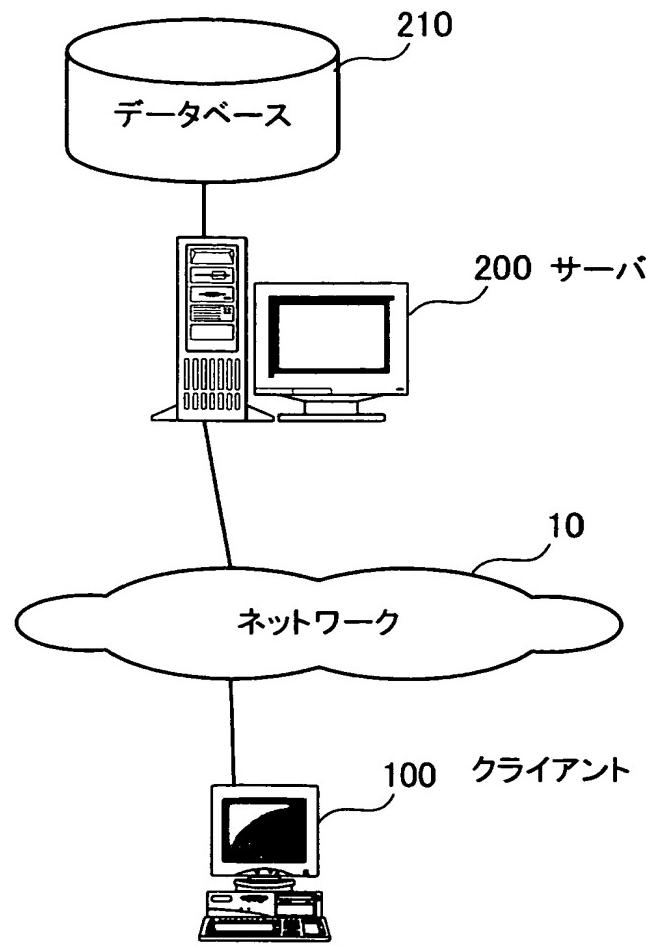
【0153】

- 1 a, 1 b, 1 c 文書
- 2 特徴素抽出手段
- 3 関連度計算手段
- 4 配置手段
- 5 関係線生成手段
- 6 表示手段
- 7 関係図
- 7 a ~ 7 g オブジェクト

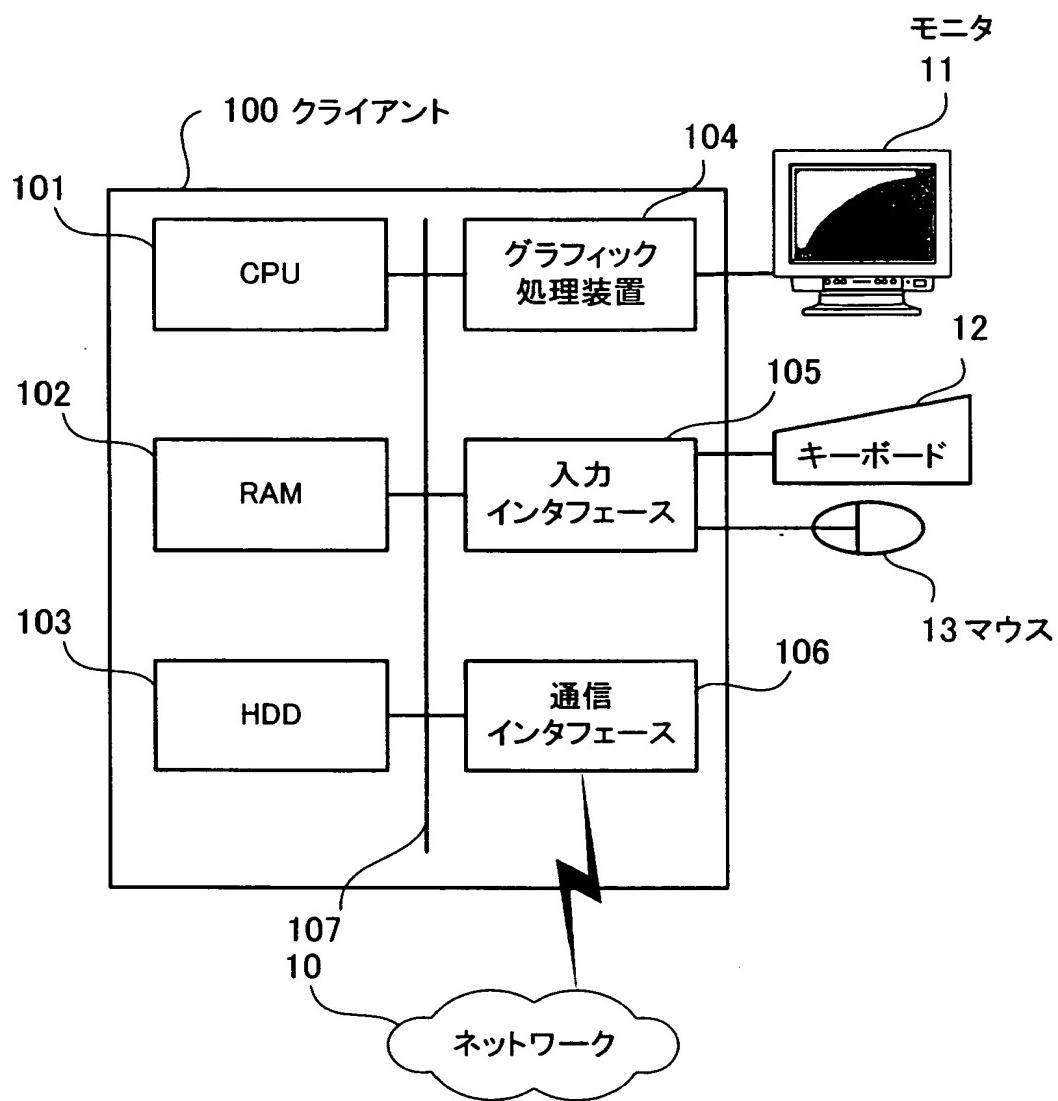
【書類名】図面
【図1】



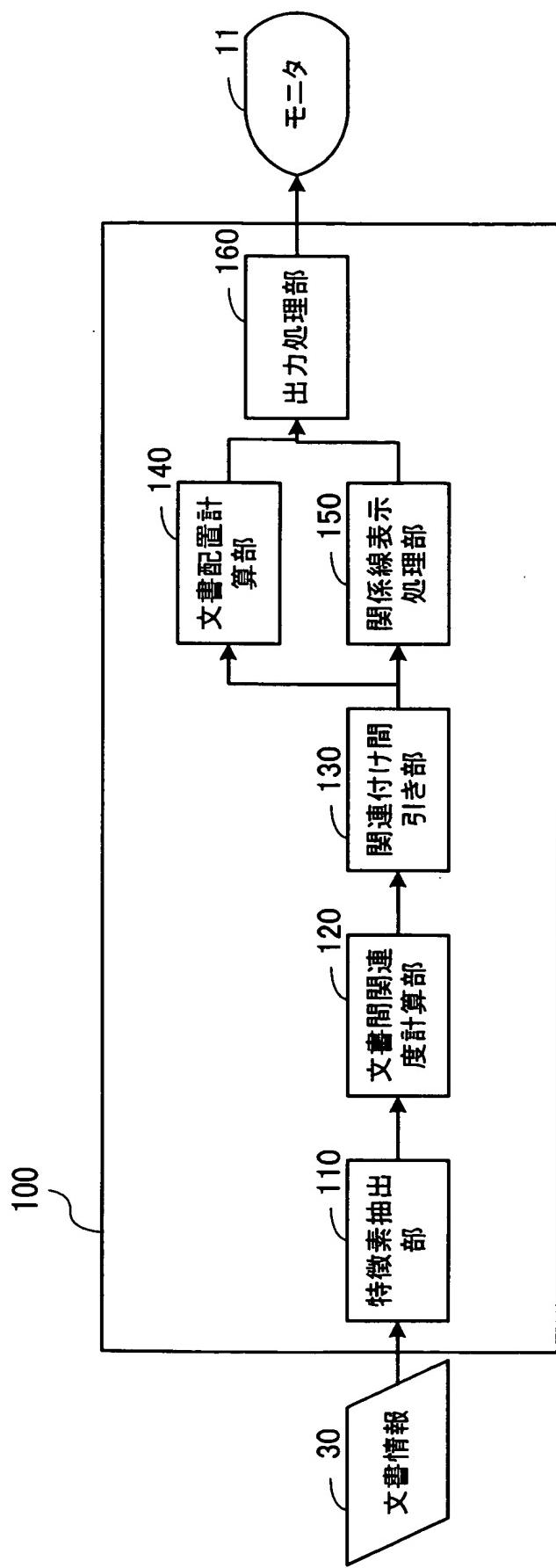
【図2】



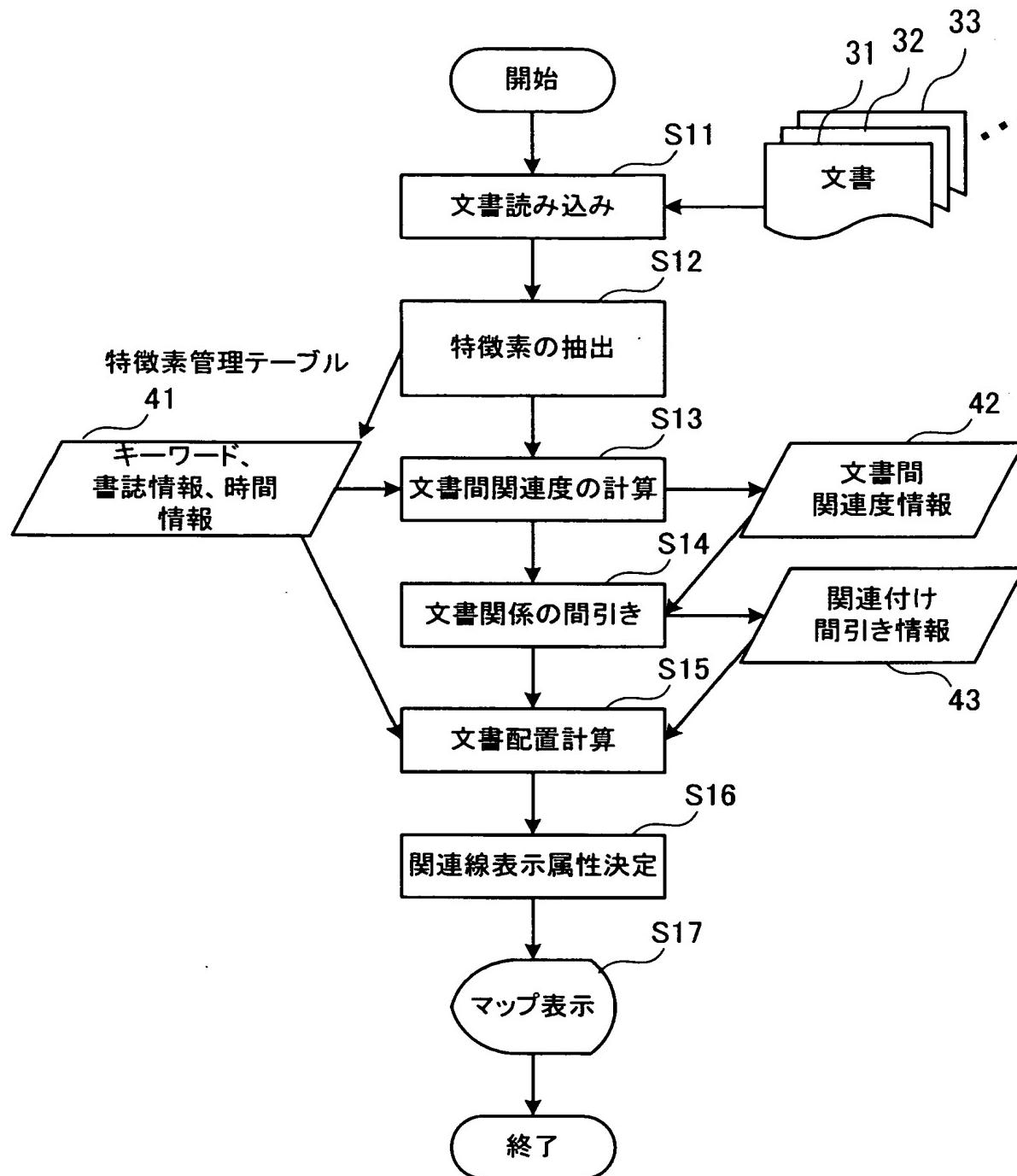
【図 3】



【図 4】



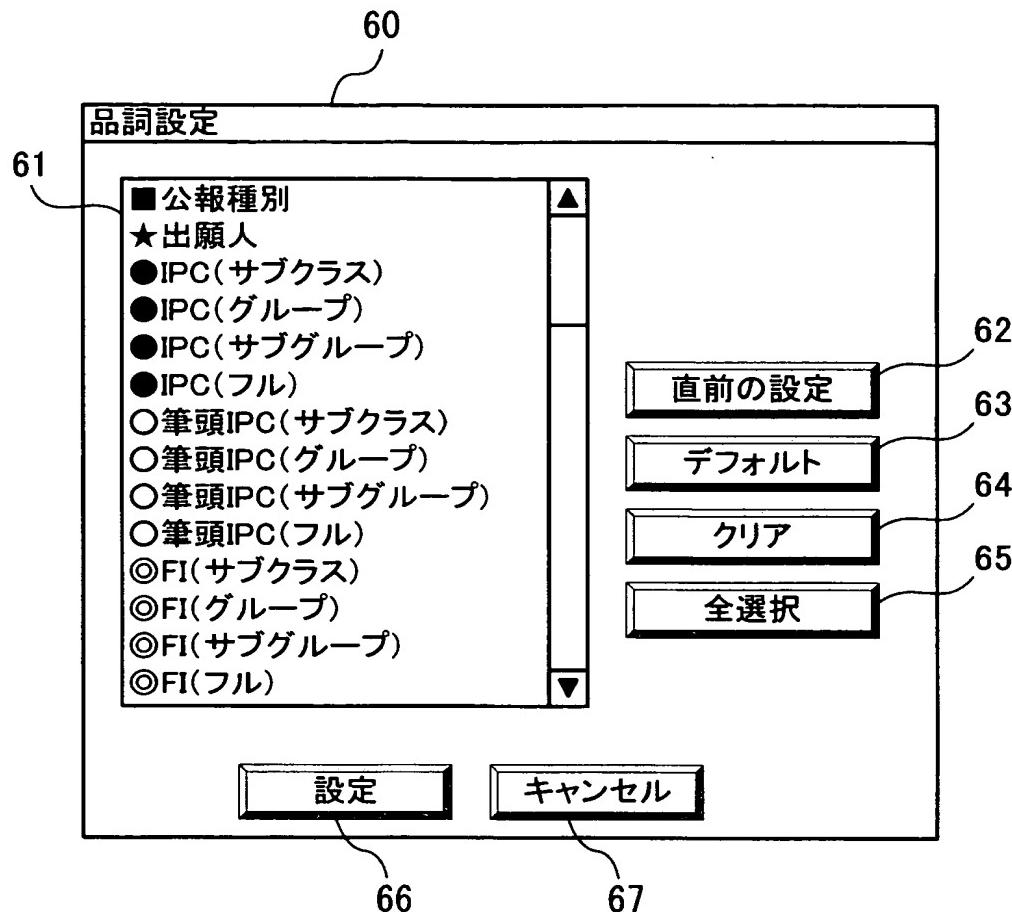
【図 5】



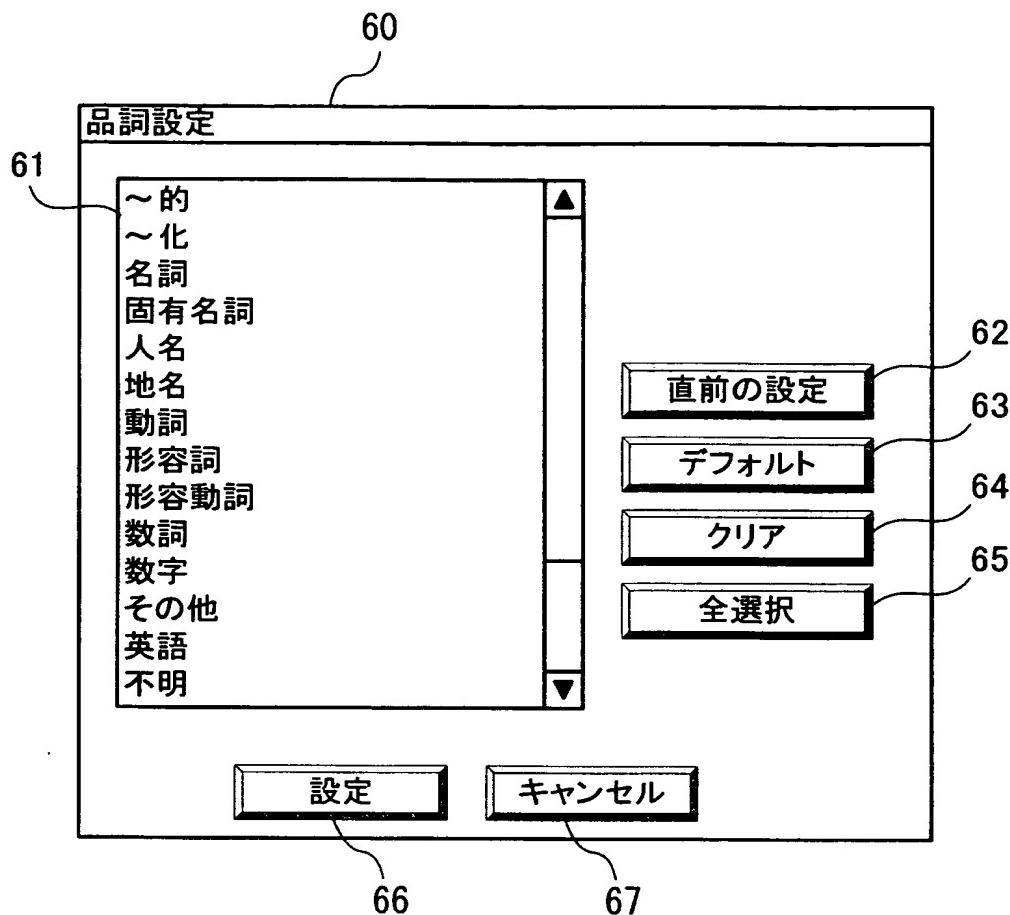
【図6】

| | | |
|--|--|---|
| 日本国特許庁(JP) 公開特許公報(A) | | 特許出願公開番号 特開2003-012345 51 公開日平成15年6月6日(2003.6.6) |
| <p>Int.Cl. 識別記号 G06F 17/60 234</p> <p>F1 G06F 17/60 234 C 5B075</p> <p>特開2003-012345 51 公開日平成15年6月6日(2003.6.6)</p> | | |
| 審査請求 未請求 請求項の数4 OL (全9頁) | | |
| 52 | <p>出願番号 特願2001-123456 出願日 平成13年11月28日</p> <p>優先権主張番号 2001-012345 優先日 平成13年2月1日</p> <p>主張国 US</p> | <p>(71)出願人 F通株式会社 神奈川県川崎市中原区○○丁目○番○号</p> <p>(72)発明者 田中一成 神奈川県川崎市中原区○○丁目○番○号</p> <p>(72)発明者 渡部勇 神奈川県川崎市中原区○○丁目○番○号</p> <p>(74)代理人 神奈川県川崎市中原区○○丁目○番○号 弁理士 服部毅</p> |
| | | |
| 【発明の名称】関係図作成装置 | | |
| <p>【要約】</p> <p>【課題】各文書を時系列で表示する。</p> <p>【手段】</p> <p>【解決手段】</p> | | |

【図 7】



【図 8】



【図 9】

41 特徴素管理テーブル

| | | |
|-------|-----|-------------|
| キーワード | 名詞 | ワークフロー |
| | 名詞 | コンピュータ |
| | ⋮ | ⋮ |
| | ⋮ | ⋮ |
| 書誌情報 | 発明者 | 田中 一成 |
| | 発明者 | 渡部 勇 |
| | 出願人 | F通 |
| | ⋮ | ⋮ |
| 時間情報 | 出願日 | 2001年11月28日 |
| | 優先日 | 2001年2月1日 |
| | 公開日 | 2003年6月6日 |
| | ⋮ | ⋮ |

【図 10】

41a 文書一単語行列

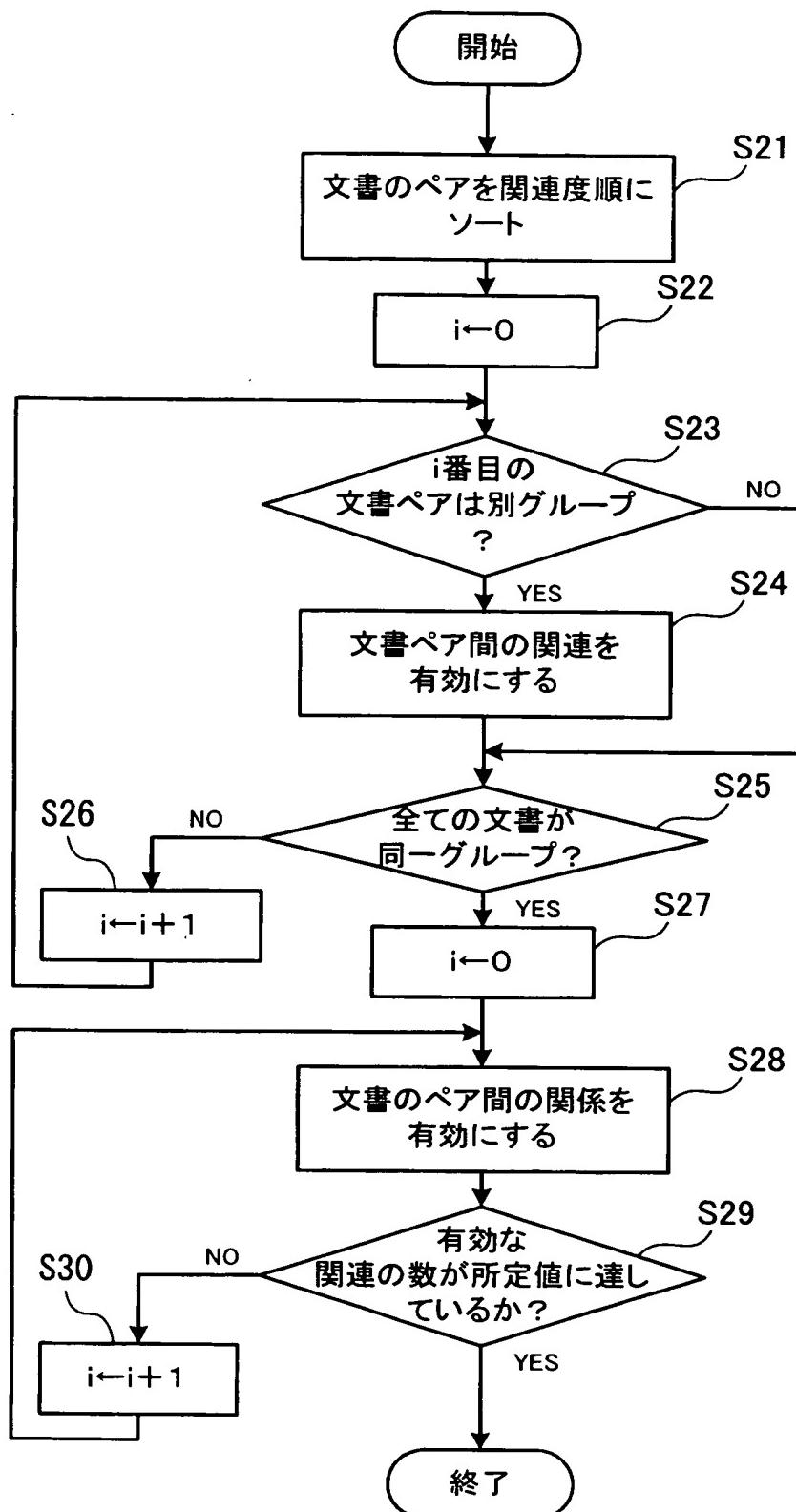
| | G06F 3/12 IPC | G06F 3/12 N:FI | グラフィカル:名詞 | 印刷:名詞 | 取り込む:動詞 | ... |
|------|---------------|----------------|-----------|-------|---------|-----|
| 文書#1 | 1 | 1 | 1 | 1 | 1 | ... |
| 文書#2 | 1 | 0 | 0 | 0 | 0 | ... |
| 文書#3 | 0 | 1 | 0 | 0 | 0 | ... |
| 文書#4 | 0 | 0 | 0 | 1 | 1 | ... |
| .. | .. | .. | .. | .. | .. | .. |

【図11】

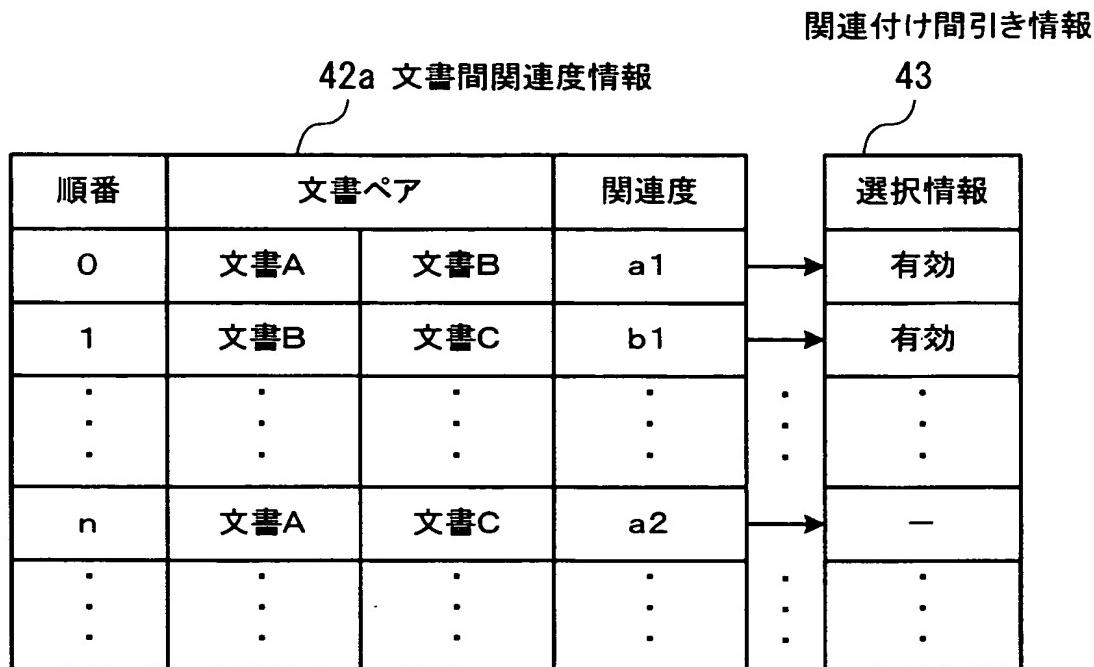
42 文書間関連度情報

| 文書ペア | | 関連度 |
|------|-----|-----|
| 文書A | 文書B | a1 |
| 文書A | 文書C | a2 |
| . | . | . |
| . | . | . |
| 文書B | 文書C | b1 |
| . | . | . |
| . | . | . |

【図 12】



【図13】



【図 1-4】

間引き設定

条件

91a エッジの数 0 %

92a 関速度 0 以下

93a 平均次数 3 に減らす

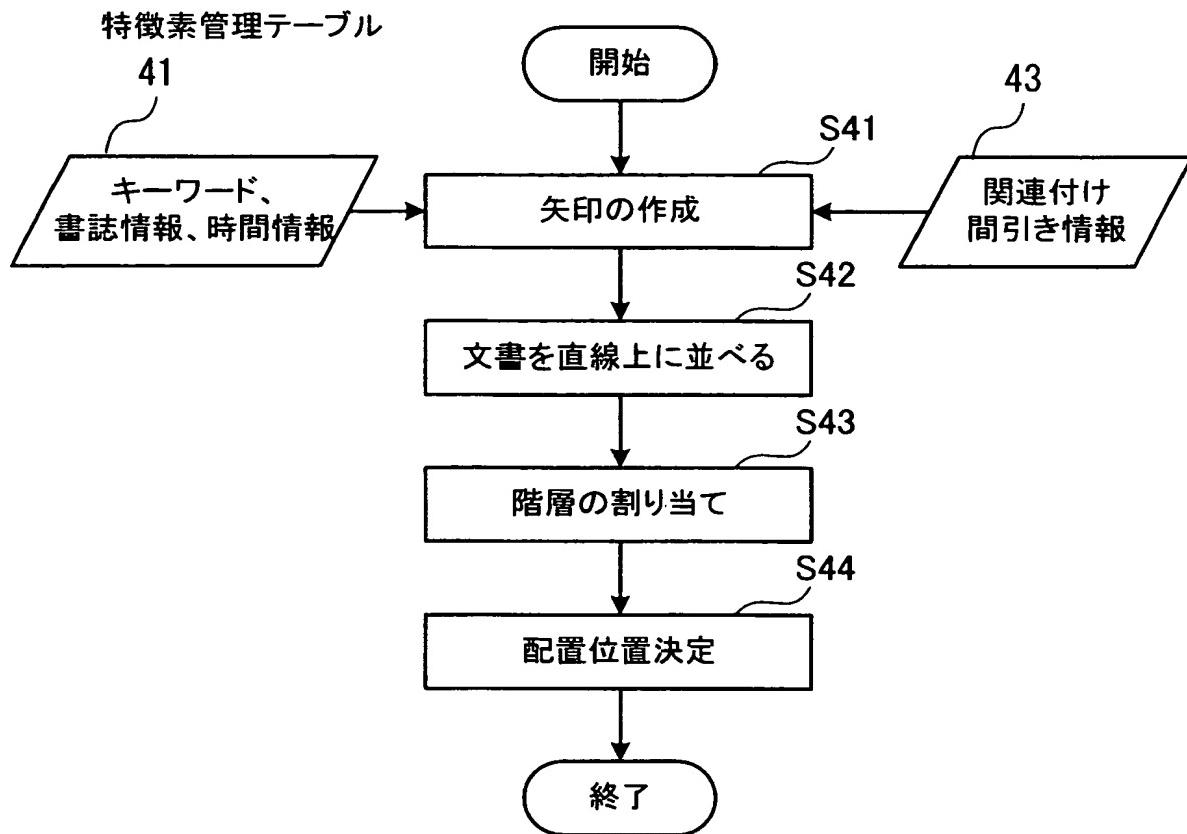
94 連結性を保持する

95 間引きしたエッジを透明化する

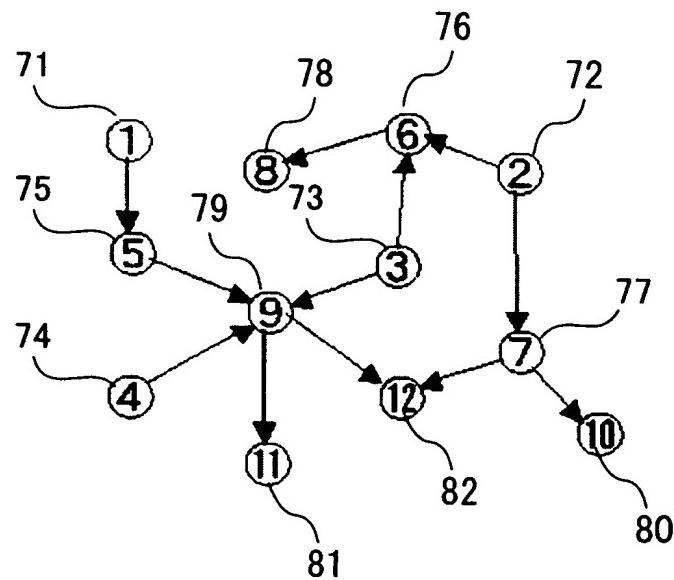
1ノード当たりの最大エッジ数 5

OK キャンセル

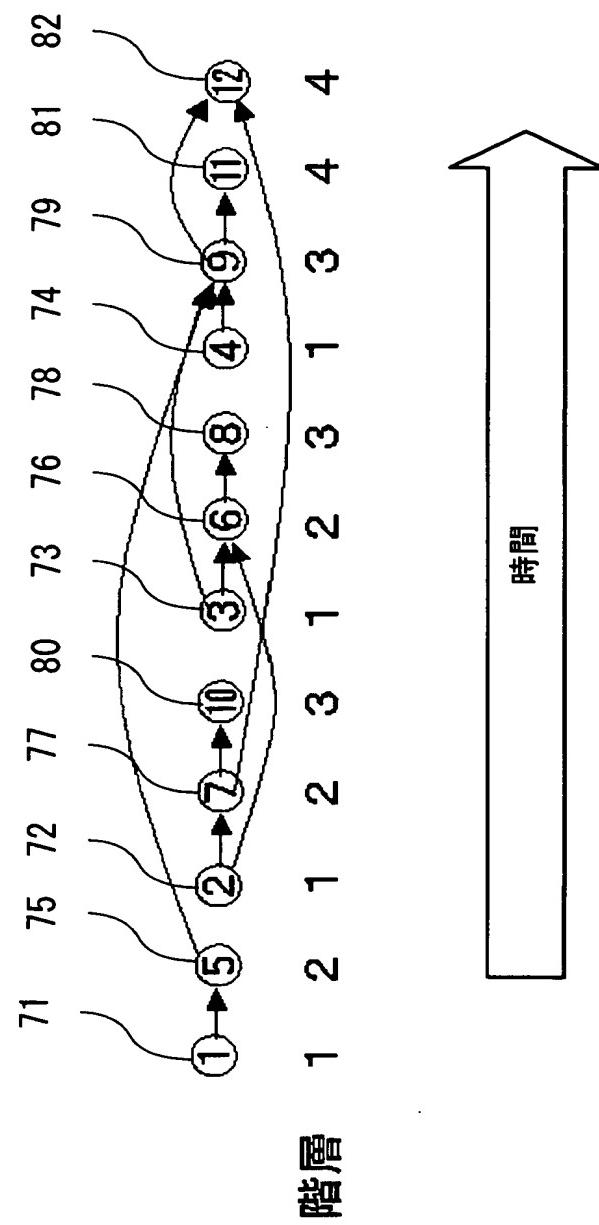
【図 15】



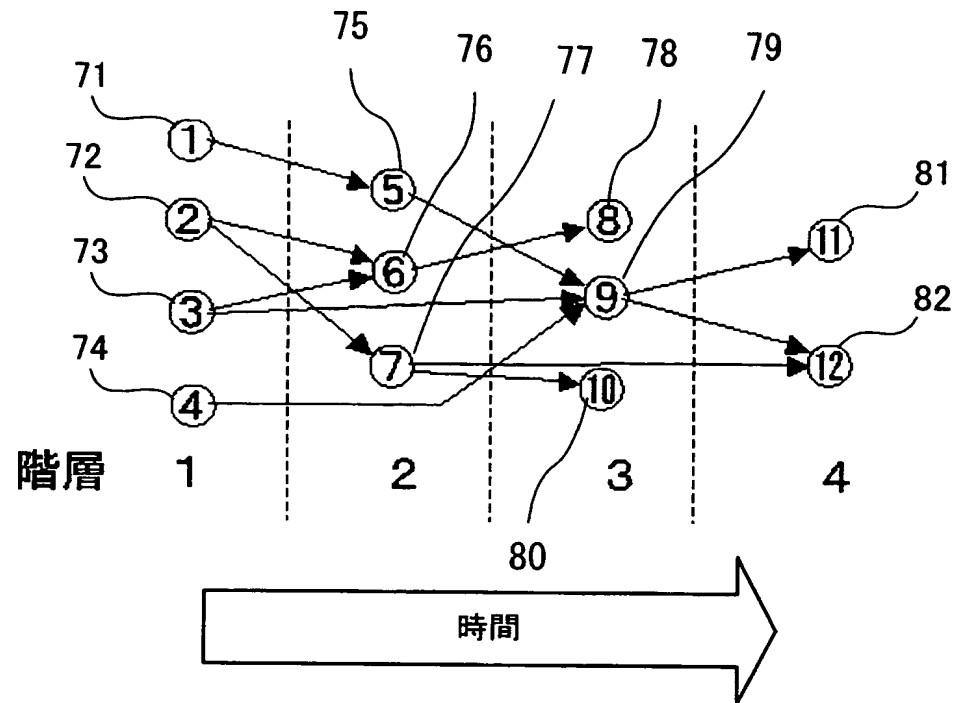
【図 16】



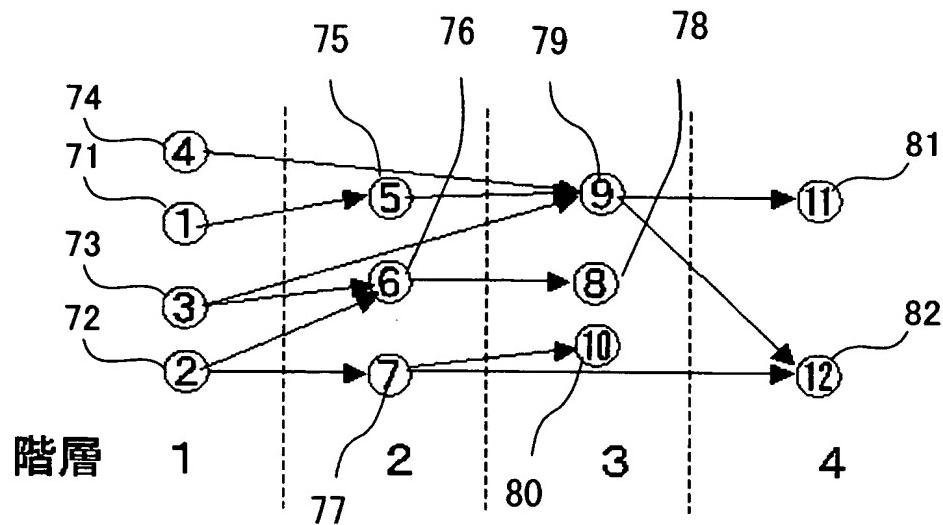
【図17】



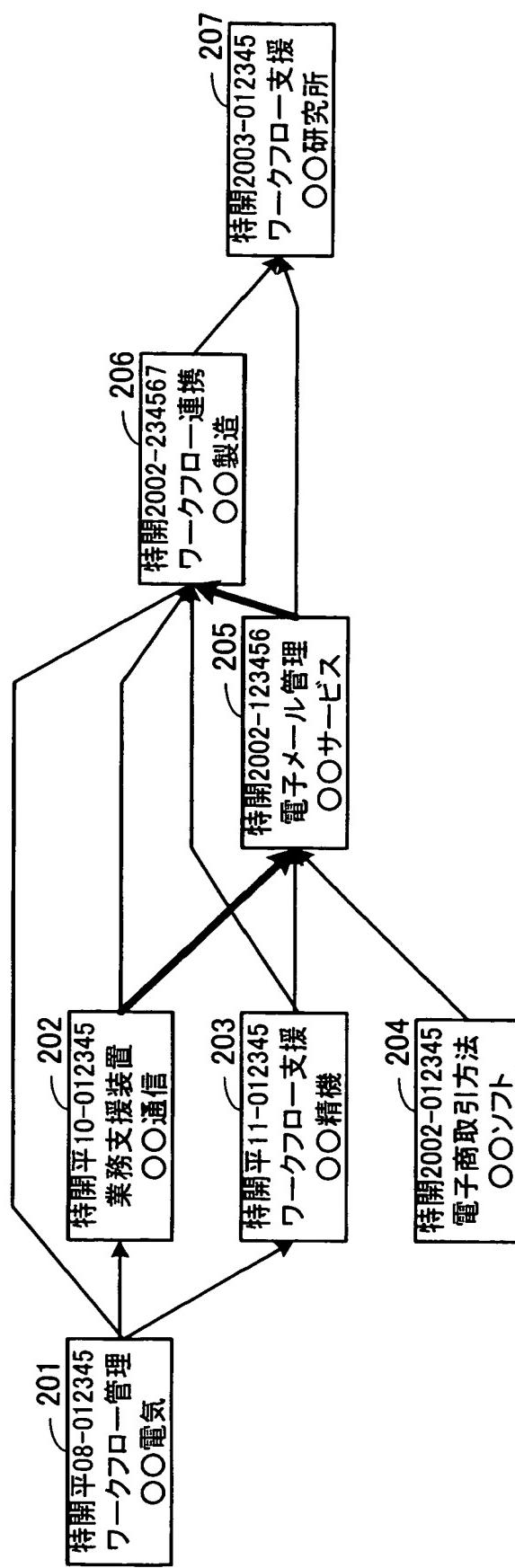
【図18】



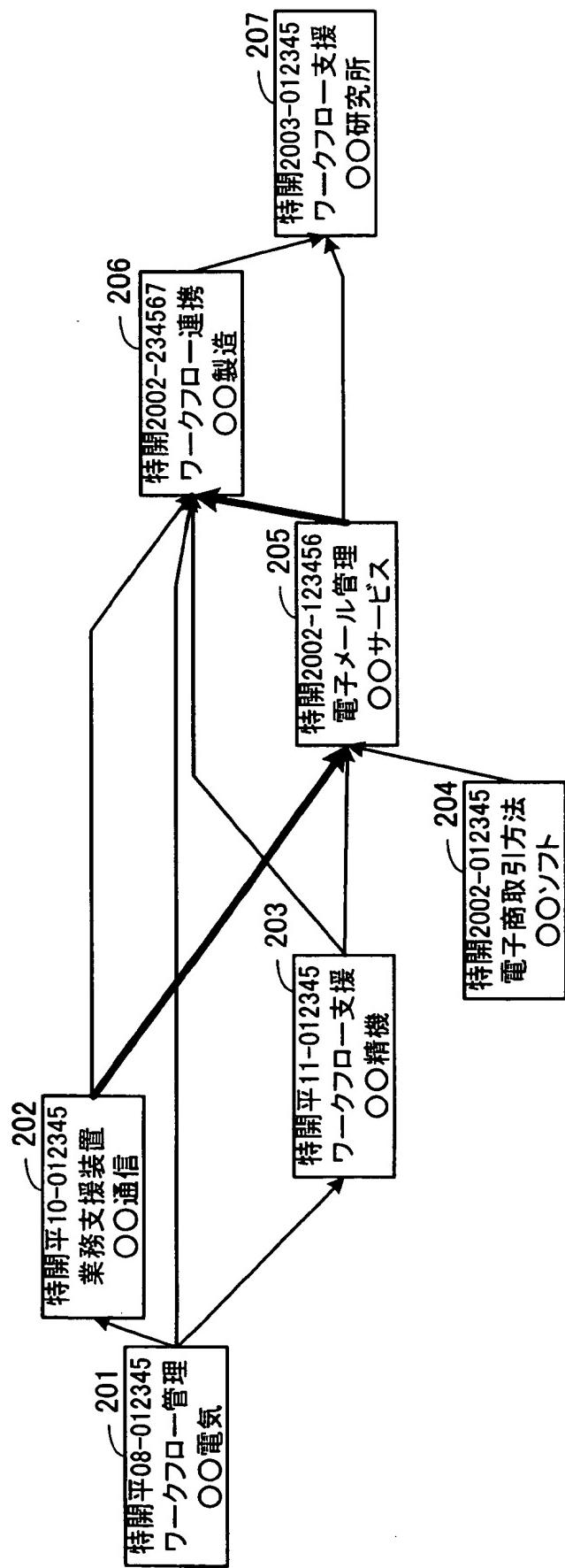
【図19】



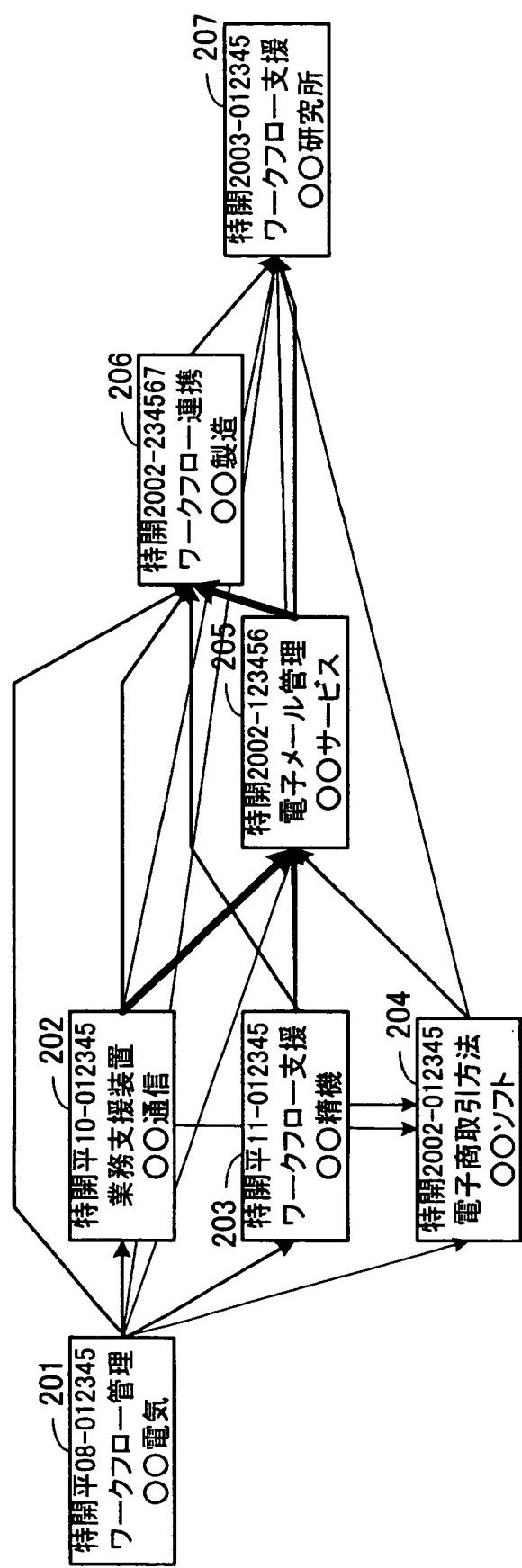
【図20】



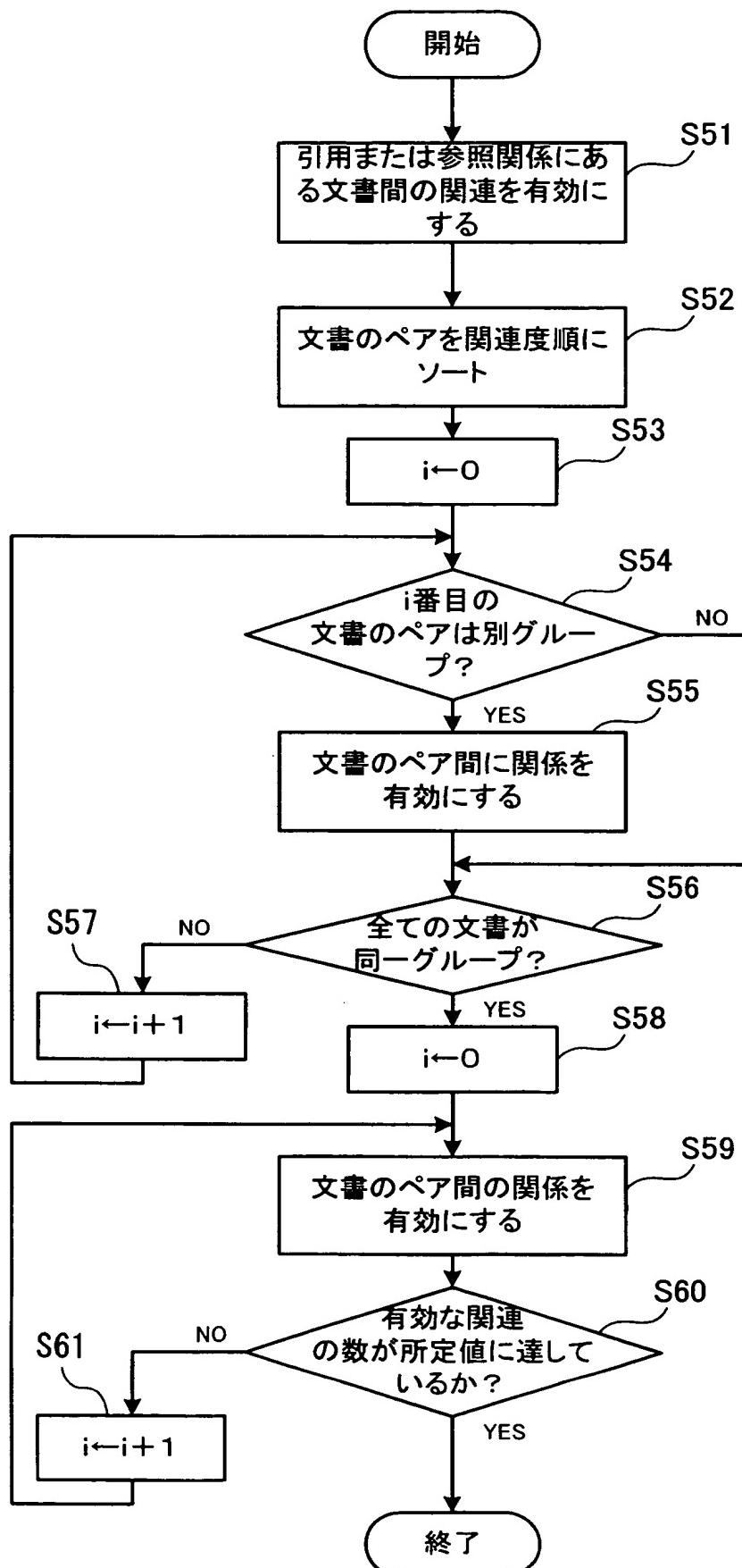
【図 21】



【図22】



【図23】





【書類名】要約書

【要約】

【課題】 引用関係や参照関係が明示されていない文書間の関連度を明確にした上で、各文書を時系列で表示できるようにする。

【解決手段】 複数の文書 1 a, 1 b, 1 c, . . . が入力されると、特徴素抽出手段 2 により、複数の文書 1 a, 1 b, 1 c, . . . から時間情報を含む特徴素が抽出される。関連度計算手段 3 により、複数の文書 1 a, 1 b, 1 c, . . . から抽出された特徴素に基づいて、文書間の関連度が計算される。さらに、配置手段 4 により、複数の文書 1 a, 1 b, 1 c, . . . から抽出された時間情報に基づいて、各文書のオブジェクト 7 a ~ 7 g が時間軸上に配置される。また、関係線生成手段 5 により、文書 1 a, 1 b, 1 c, . . . 間の関連度に基づいてオブジェクト間を結ぶ関係線が生成される。そして、表示手段 6 により、各文書のオブジェクト 7 a ~ 7 g と関係線とで構成される関係図 7 が表示される。

【選択図】 図 1

特願2003-353928

出願人履歴情報

識別番号 [000005223]

1. 変更年月日 1996年 3月26日
[変更理由] 住所変更
住 所 神奈川県川崎市中原区上小田中4丁目1番1号
氏 名 富士通株式会社